COMPUTER PROGRAMS FOR SCORING TEST DATA WITH ITEM CHARACTERISTIC CURVE MODELS

Isaac I. Bejar
and
David J. Weiss

RESEARCH REPORT 79-1
FEBRUARY 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Army Research Institute, and Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Computer Programs for Scoring Test Data with Item Characteristic Curve Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Isaac I. Bejar and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.: NR150-389
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE February 1979 13. NUMBER OF PAGES 84
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 in 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

This research was supported by funds from the Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Army Research Institute, and the Office of Naval Research, and monitored by the Office of Naval Research.

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

item characteristic curve theory	test theory	psychological testing
ICC	test scoring	adaptive testing
item response theory	ability testing	computerized testing
latent trait theory	achievement testing	tailored testing

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

Three computer programs are described for scoring test response data using item characteristic curve (ICC) or latent trait models. The rationale and mathematical basis of both maximum likelihood and Bayesian ICC scoring methods are presented, as well as some data comparing the two methods of scoring. The three computer programs are designed for scoring conventional (linear) test data (LINDSCO) in dichotomous response format, adaptive test dichotomous data (ADADSCO), and conventional (linear) test data scored by polychotomous ICC models (LINPSCO). Options available in these three general

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)*		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)*		8. CONTRACT OR GRANT NUMBER*(s)*
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)*
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)*		
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)*		
20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*		

purpose programs are described, and examples of the input and output are given for each program. Complete FORTRAN listings of the three programs are included.

CONTENTS

Appendices

Acknowledgments

The computer programs described in this report were written by Sinan N. Neftci, Meera Mahesh Pondicherry, and Ann Marie Kohler. Their contributions are gratefully acknowledged. The assistance of James B. Sympson in preparing the section entitled ICC-Based Scoring is also appreciated.

Disclaimer

While every attempt has been made to insure the accuracy of the computer programs described in this report, the authors assume no responsibility for the accuracy or performance of the programs.

Technical Editor: Barbara Leslie Camm

Computer Programs for Scoring Test Data with Item Characteristic Curve Models

Although latent trait test theory, or item characteristic curve (ICC) theory, has been developing since Lawley's (1943) paper more than 30 years ago, applications of the theory have appeared only recently. However, there are indications that latent trait test theory is beginning to reach the practitioner who is concerned with test development and usage in applied settings. This is evidenced,not only by the increasing number of journal articles concerned with latent trait test theory (e.g., the summer 1977 special issue of the Journal of Educational Measurement on applications of latent trait models) and in presentations and training sessions at professional meetings, but also by its application in adaptive (Weiss, 1976) or tailored (Lord, 1970) testing.

A potential disadvantage of latent trait test theory is that its use often involves complex computational procedures. To apply ICC models to the development of tests and their scoring, the psychometrician must be able to estimate the ICC parameters of the items in the test,and then use them in conjunction with the response data of a new group of testees in order to estimate their trait scores (e.g., ability or achievement levels). A number of computer programs are available for estimating ICC item parameters (these are summarized in Appendix Table A). However, there appeared to be no general programs available for scoring test data with ICC models when item parameter estimates were available from previous data sets. This report describes several programs designed to meet this need.

An Introduction to Test Scoring

The problem of test scoring can be conceptualized as the process of summarizing a testee's answers to a set of test questions into a single number in such a way that the score will be indicative of the testee's position on the trait being measured by the test. The most common test scoring strategy is to add the number of correct answers and to transform the score into some type of standard score or percentile to add interpretability. Historically, the number-correct score has been used because it is easy to calculate, and in pre-computer days this was an essential requirement of a test scoring procedure. As a general procedure for scoring tests of ability and achievement, however, the number-correct score has several deficiencies.

Inadequacies of the Number-Correct Score

One major problem with the number-correct score is that it is possible for the same number-correct score to be obtained in several different ways; that is, several response patterns can result in the same number-correct score. If the items in a test are all of equal difficulty and discrimination, and therefore are essentially replicates of each other, this will have little effect on the number-correct score, since different response patterns among

replicate items are of little consequence. But it is a very rare test--and one which would have little general measurement utility--which would have items that are all replicates of each other with regard to difficulty and discrimination.

When test items differ with respect to difficulty or discrimination, they are no longer replicates. Under these circumstances, different patterns of response to the same set of items convey different information with regard to a testee's trait level. The testee who correctly answers only five very difficult items in a test is likely of higher ability than the testee who correctly answers only five very easy items in the same test. Although the total number-correct score is the same for these two testees, their trait level estimates derived from latent trait or ICC theory will differ. An additional unattractive feature of the number-correct score is the fact that the number of possible scores is determined by the number of items in the test. Thus, if a test consists of only 10 items, only 10 unique scores are possible. Although this may be sufficient in some applications, in others it might be desirable to obtain a finer gradation of scores.

The inadequacy of the number-correct score as a general test-scoring procedure is most obvious when considering how to score responses of testees who have been administered different sets of items, as in adaptive or tailored testing. In these kinds of tests, number-correct scores are completely inappropriate, since different testees will receive items of different difficulties and discriminations as well as different numbers of items in an adaptive test. In addition, the proportion of correct responses obtained by all testees will be approximately the same in a well-designed adaptive test (e.g., Weiss, 1975).

ICC-Based Scoring

The scoring programs described in this report use considerably more refined approaches than a mere adding of correct answers and are usable for scoring both conventional and adaptive test data. This refinement is possible because ICC theory makes very explicit specifications about the relationship between performance on a test item and the testee's position on the trait, θ. This relationship is referred to as the *item characteristic curve* (ICC; Lord & Novick, 1968) when the items are scored into two categories (correct or incorrect) or, when there are more than two score categories, as the *operating characteristic function* (Samejima, 1969).

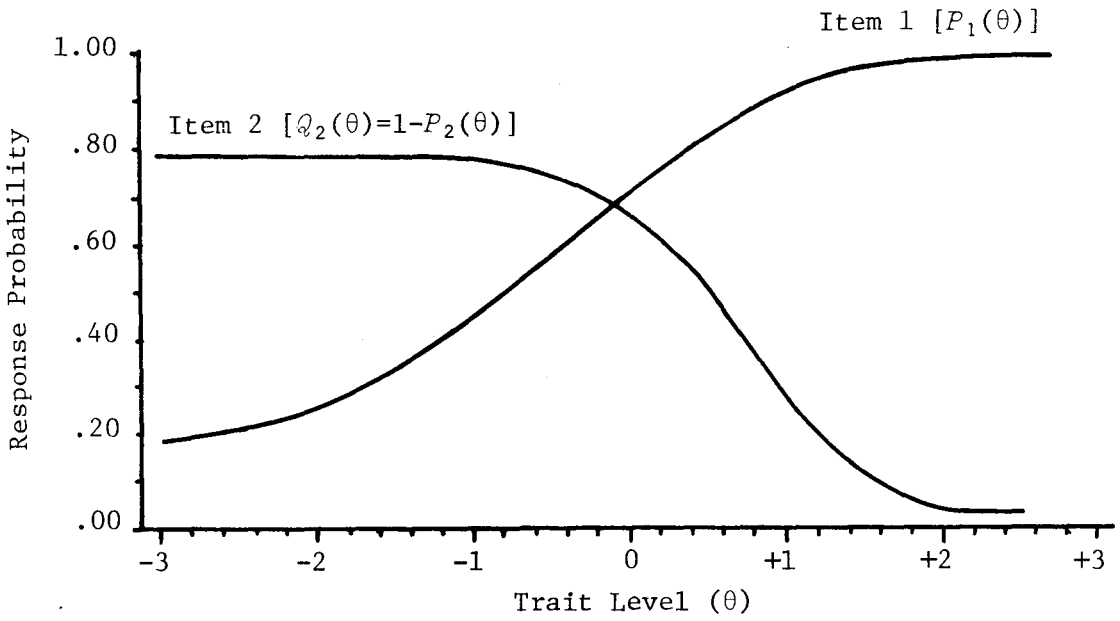
In the context of latent trait test theory, scoring may be conceptualized as finding the value of θ (i.e., the trait being measured) most "compatible," in some sense, with a given pattern of responses to the test items, given the ICC item parameters for each item answered. For maximum likelihood scoring, the score associated with a given response vector is that value of θ for which the likelihood of the response vector is maximum. For Bayesian scoring, the score is usually either the value of θ that minimizes the mean squared difference between the estimated θ and the "true" θ, or the value of θ that is most probable given the observed responses.

Maximum likelihood scoring. The details of the maximum likelihood scoring procedure are presented below. However, a conceptual explanation based on two

dichotomously scored test items will serve to explicate its rationale.

Figure 1 shows response probability curves for two test items--Item 1, which was answered correctly (resulting in an ICC plot of the probability of a correct response), and Item 2, which was answered incorrectly (resulting in a descending plot of the probability of an incorrect response, or 1 minus the ICC). The ICC curves for the two items are described by three parameters: (1) difficulty, b, which is the location of the ICC on the trait (θ) continuum at the point of maximum slope of the ICC (b=-.5 for Item 1 and .75 for Item 2); (2) their discrimination, a, which is proportional to the slope of the ICC at b (a=.8 for Item 1 and 1.4 for Item 2); and (3) "guessing," c, the lower asymptote of the probability of a correct response at $\theta=-\infty$ (c=.16 for Item 1 and 1-.78=.22 for Item 2).

Figure 1
Response Probability Plots for a Correctly Answered Item (Item 1) and an Incorrectly Answered Item (Item 2)



The first step in maximum likelihood scoring consists of determining the likelihood of the response pattern (correct response to Item 1 and incorrect response to Item 2). Assuming local independence, which means that responses to the test items have nothing in common except their relationship to the underlying trait, θ, the likelihood of a response pattern at any value of θ can be determined by multiplying the separate probabilities of the responses in the response pattern for that value of θ. The value of θ for which the likelihood is maximum is the maximum likelihood estimate of θ.

Conceptually, this can be illustrated with the ICCs in Figure 1 by using discrete values of θ, such as those shown in Table 1. For example, at θ=-1.0, the probability of a correct response to Item 1 (scored as 1) is .442 and the

probability of an incorrect response to Item 2 (scored as 0) is .768; multiplying these values gives the likelihood of the [1,0] response pattern as .340. At θ=+1.0, the probability of a correct response [1] to Item 1 is .903 and the probability of an incorrect response to Item 2 is .277; the likelihood of the [1,0] response pattern is therefore .250. Similarly, at θ=0.0, the probability of a correct response to Item 1 is .718 and the probability of an incorrect response to Item 2 is .668, resulting in a likelihood for the [1,0] response pattern of .479. This process of computing likelihoods for the [1,0] response pattern can be repeated for a large number of values along the θ continuum.

Table 1
Probability of a Correct Response to Item 1 [$P_1(\theta)$] and Probability of an Incorrect Response to Item 2 [$Q_2(\theta)$] for Selected Values of θ (Item 1: a=.8, b=-.5, c=.16; Item 2: a=1.4, b=.75, c=.22), and Values of the Likelihood Function [$L(\theta)$]

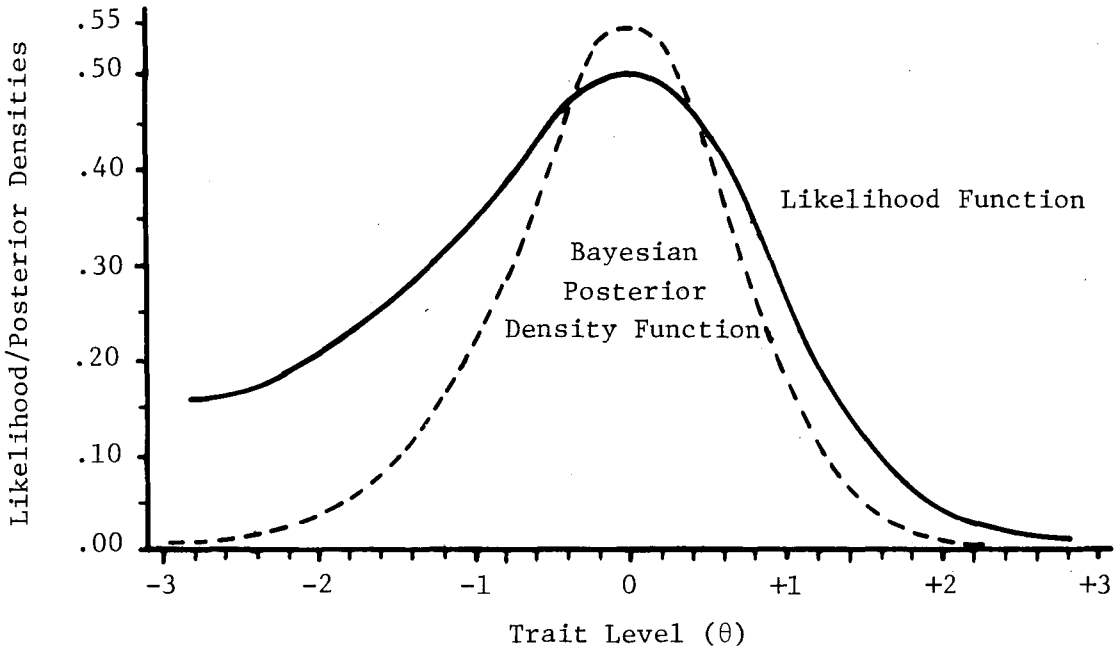
θ	$P_1(\theta)$	$Q_2(\theta)$	$L(\theta)$
-3.0	.187	.780	.146
-2.5	.212	.780	.165
-2.0	.257	.779	.200
-1.5	.332	.776	.257
-1.0	.442	.768	.340
.5	.580	.742	.430
0.0	.718	.668	.479
.5	.828	.503	.416
1.0	.903	.277	.250
1.5	.948	.112	.106
2.0	.973	.038	.037
2.5	.986	.012	.012
3.0	.993	.004	.004

The result of computing likelihoods for all possible values of θ based on the response pattern and the relevant ICCs can be a plot of the likelihood values as a function of θ. This plot, shown as a solid curve in Figure 2, is called a likelihood function. As can be seen, the maximum of the likelihood function in Figure 2 occurs at about θ=0.0 (actually .01). Thus, θ=.01 can be considered the maximum likelihood estimate of θ associated with the [1,0] response pattern, given the parameters of the ICCs for the items generating that response pattern. The maximum likelihood θ estimate is thus the value of θ which maximizes likelihood of the given response pattern for items with the specified ICCs.

The generalization of the scoring method for more than two items is straightforward. For each value of θ, the likelihood would be determined by multiplying the response probabilities for the appropriate ICCs (based on the specified response pattern) across all items that have been answered. Thus, for n items, n probabilities would be multiplied at each value of θ to obtain the likelihoods. The resultant likelihood values for all values of θ

could be plotted; and the maximum of the likelihood function would be used to identify the value of θ that gives the observed response pattern the greatest probability of occurrence.

Figure 2
Likelihood Function and Bayesian Posterior Density Function for the [1,0] Response Pattern



Maximum likelihood scores are intuitively appealing; at the same time, they have a number of optimal statistical characteristics, at least asymptotically (i.e., when large numbers of items are administered). Of special relevance is the fact that as the number of items in the response pattern increases, it can be shown (Kendall & Stuart, 1961) that maximum likelihood estimates have minimum variance; and the reciprocal of that variance is known as the information function of θ. As a consequence, different scores (i.e., θ estimates) can have different degrees of accuracy as estimators of θ (Birnbaum, 1968; Samejima, 1969).

Bayesian scoring. Although the numerical details of maximum likelihood and Bayesian scoring are substantially different, the two methods are conceptually very similar. Bayesian scores are based on the likelihood function modified by the prior probability density function of θ. The prior probability density function describes the assumed distribution of θ in the population of individuals to be tested.

To illustrate, call $L(\theta)$ the likelihood of the response pattern for a given θ value. Now call $f(\theta)$ the prior probability density associated with that value of θ. The modified likelihood, which may be called $p(\theta)$ is then

$$p(\theta)=f(\theta)L(\theta)/\int[f(\theta)L(\theta)]d\theta. \qquad [1]$$

Equation 1 is called the posterior probability density function. Just as the maximum likelihood score is the value of θ for which $L(\theta)$ is maximum, one kind of Bayesian score is the value of θ for which $p(\theta)$ is maximum. Such scores are called Bayes modal estimates by Samejima (1969) because they are based on the mode of the posterior density function. A different type of Bayesian estimate is based on the mean of the posterior density function. Owen's (1975) Bayesian scoring procedure, which will be described in detail below, is an example of this approach. In his procedure, the prior probability densities are provided by a normal density function. Other Bayesian scoring procedures are also available (Sympson, 1977).

These concepts can be illustrated using the likelihoods associated with the [1,0] response pattern discussed earlier. Table 2 shows for several values of θ the probability of a correct response to the first item $[P_1(\theta)]$; the probability of an incorrect response to the second item $[Q_2(\theta)]$; the likelihood of the response pattern $[L(\theta)]$ (these first three columns correspond to the data in Table 1); the prior probability densities $[f(\theta)]$, which in this case are ordinates of a normal distribution with mean of zero and standard deviation of 1; and the posterior density function $[p(\theta)]$, computed using Equation 1. For these data

$$\int [f(\theta)L(\theta)]d\theta \simeq .348. \qquad [2]$$

The resulting posterior density function, $[p(\theta)]$, is shown as the dashed curve in Figure 2.

Table 2
Response Probabilities $[P_1(\theta), Q_2(\theta)]$, Likelihoods $[L(\theta)]$, Weights $[w(\theta)]$, and Posterior Density Function $[p(\theta)]$ for a Two-Item Response Pattern

θ	$P_1(\theta)$	$Q_2(\theta)$	$L(\theta)$	$f(\theta)$	$p(\theta)$
-3.0	.187	.780	.146	.004	.002
-2.5	.212	.780	.165	.018	.009
-2.0	.257	.779	.200	.054	.031
-1.5	.332	.776	.257	.130	.096
-1.0	.442	.768	.340	.242	.236
-0.5	.580	.742	.430	.352	.435
0.0	.718	.668	.479	.399	.549
0.5	.828	.503	.416	.352	.421
1.0	.903	.277	.250	.242	.174
1.5	.948	.112	.106	.130	.040
2.0	.973	.038	.037	.054	.006
2.5	.986	.012	.012	.018	.001
3.0	.993	.004	.004	.004	.000

The mode of the posterior density function in Figure 2 is located near $\theta=0$, so the Bayesian modal estimate and the maximum likelihood estimate are about the same for this data. The Bayesian θ estimate based on the mean of the $p(\theta)$ distribution is -.12. This θ estimate does not coincide with the maximum likelihood estimate ($\hat{\theta}=.01$); as will be further shown below, estimates of θ obtained from different ICC scoring methods do not generally agree.

Differences Among Scoring Methods

The programs described in this report are capable of scoring test data using most of the ICC response models available. The selection among models should not be arbitrary, especially when individual decisions are to be made on the basis of test scores. Dichotomous data can be scored by means of the one-, two-, and three-parameter ICC models, using either a normal or logistic ogive ICC. Thus, given the decision with regard to the number of parameters that describe the ICC, there still remains the problem of choosing between the normal or logistic ogive response models for scoring purposes. Unfortunately, there are as yet no firm guidelines for choosing between these two response models. Samejima (1969) has shown that the normal and logistic ogive models differ with respect to their scoring "philosophies," but the practical implications of these differences remain to be investigated.

To illustrate the differences among the models and different ICC scoring procedures, all response patterns for a five-item test were scored by maximum likelihood, assuming both normal and logistic ogive ICCs, and by Owen's (1975) Bayesian scoring method. Table 3 gives the item parameters assumed for the hypothetical five-item test. For all items, the c (guessing) parameter was set at 0.0, indicating that a two-parameter ICC model was used. Items varied in difficulty (b) from -2 to +2 and had discriminations of 1.00 or 1.50.

Table 3
Item Paremeters for Five-Item Test

Item	a	b	c
1	1.00	-2.00	.00
2	1.50	-1.00	.00
3	1.00	0.00	.00
4	1.50	1.00	.00
5	1.00	2.00	.00

In a five-item test in which each item is scored dichotomously, there are $2^5=32$ different response patterns. These response patterns are shown in Table 4 along with the scores associated with them. It is obvious from the data in Table 4 that for a given response pattern, the scores (all of which are on the same metric) differed somewhat. This indicates that the scoring procedures are not interchangeable.

For example, consider the five response patterns which have 20% correct, namely Patterns 2, 3, 5, 9, and 17. Not only do the θ estimates (scores) for a given response pattern differ among the three scoring procedures, but there are some differences in the ordering of the θ estimates derived from these response patterns within each procedure. For maximum likelihood scoring using

a normal ogive ICC, the ordering of the θ estimates derived from the five response patterns was exactly the same as that obtained from the Bayesian scoring procedure, although the numerical values of the θ estimates were uniformly higher for the Bayesian procedure. For both these scoring methods, there was a tendency for higher ability estimates to be obtained when a more difficult item was answered correctly. For example, the lowest θ estimate was obtained by both scoring methods when the easiest item (Item 1) was answered correctly (Response Pattern 17); when only Item 2 was answered correctly (Response Pattern 9), the θ estimates from both the Bayesian and maximum likelihood normal procedures increased. In addition, both scoring methods took into account the discriminations of the items involved. For example, Response

Table 4
Scores Given to Each Response Pattern by Three Scoring Methods

Response Pattern	Maximum Likelihood		Bayesian
	Normal	Logistic	
1. 00000	∞*	∞*	-1.72
2. 00001	-.93	-1.60	-.64
3. 00010	-.61	-1.19	-.38
4. 00011	-.13	-.46	.11
5. 00100	-1.42	-1.60	-1.06
6. 00101	-.50	-.84	-.28
7. 00110	-.30	-.46	-.11
8. 00111	.13	.46	.30
9. 01000	-1.24	-1.19	-.89
10. 01001	-.23	-.46	-.15
11. 01010	.03	.00	.00
12. 01011	.50	.84	.41
13. 01100	-.60	-.46	-.42
14. 01101	.23	.46	.17
15. 01110	.39	.84	.28
16. 01111	.93	1.60	.64
17. 10000	-1.63	-1.60	-1.16
18. 10001	-.39	-.84	-.24
19. 10010	-.17	-.46	-.06
20. 10011	.30	.46	.39
21. 10100	-.78	-.84	-.58
22. 10101	.03	.00	.11
23. 10110	.17	.46	.23
24. 10111	.61	1.19	.62
25. 11000	-.42	-.46	-.29
26. 11001	.60	.46	.51
27. 11010	.78	.84	.63
28. 11011	1.42	1.60	1.09
29. 11100	.42	.46	.31
30. 11101	1.24	1.19	.93
31. 11110	1.63	1.60	1.08
32. 11111	∞*	∞*	1.55

*For maximum likelihood scoring, it is not possible to score response patterns with all correct or incorrect answers.

Pattern 2 (with a correct response to Item 5, the most difficult item) was assigned higher scores than Pattern 5; but Pattern 2 was assigned lower scores than Pattern 3 (which had a correct response to Item 4, the second most difficult item), since in Response Pattern 3 a correct answer was given to an item (Item 4) with a higher discrimination than that of Pattern 2 (Item 5).

On the other hand, assuming a logistic ogive ICC for the maximum likelihood scoring procedure, estimated values of θ were related to the discriminations of the items answered correctly. Those response patterns for which the discriminations of the items answered correctly were the same were assigned the same score, namely -1.60 for Patterns 2, 5, and 17 and -1.19 for Patterns 3 and 9. For the latter two response patterns, the discriminations of the items answered correctly were 1.50; for the former three response patterns, they were 1.00. Thus, the magnitude of the scores was a function of the item discriminations, and the item difficulties did not affect the θ estimates.

These data indicate that the assumption of different forms of the ICC within the maximum likelihood scoring procedure will, in general, result in different θ estimates. Since the Bayesian θ estimates were different from both the maximum likelihood estimates, these three ICC-based scoring procedures are not interchangeable. However, additional research is required to further delineate the similarities and differences among the θ estimates derived by different ICC-based scoring procedures and, more importantly, to assess the implications of these differences in practical applications.

General Description of the Programs

This report describes three computer programs for scoring test data with ICC models--LINDSCO, ADADSCO, and LINPSCO. Table 5 summarizes the major features of these programs. LINDSCO (LINear Dichotomous SCOring) is designed

Table 5
Summary of Program Capabilities

Model and Scoring Procedure	Dichotomous		Polychotomous (LINPSCO)	
	Linear (LINDSCO)	Adaptive (ADADSCO)	Graded	Nominal
Logistic Ogive				
Bayesian[a]	NO	NO	NO	NO
Maximum Likelihood	YES	YES	YES	YES
Normal Ogive				
Bayesian[a]	YES	YES	NO	NO
Maximum Likelihood	YES	YES	YES	NO

[a]The Bayesian scoring procedure is based on Owen (1975).

to be used for scoring test data for conventional (linear) tests in which all items are administered to each testee. It requires responses to be dichotomous; that is, responses are scored into one of two categories, such as "correct"

and "incorrect." Omissions are permitted, but they are ignored in the computations. The number of omitted items is tallied from the number of items administered and reported as part of the output for each testee. Either the normal or logistic ogive response model can be used with ICCs described by one, two, or three parameters for maximum likelihood scoring. Response patterns may also be scored by Owen's (1975) Bayesian method which assumes a normal ogive ICC. The user can also specify, in addition to a total test score, subscores on as many as 25 subscales.

ADADSCO (ADAptive Dichotomous SCOring) is similar to LINDSCO, but it is designed specifically for scoring item response data derived from adaptive testing. Since in adaptive testing each respondent answers a different set of test items, the program must locate for each testee the item parameter estimates of each attempted item; LINDSCO, in contrast, does the item search only once. ADADSCO also differs from LINDSCO in that it has no subscale scoring capabilities.

LINPSCO (LINear Polychotomous SCOring) is designed to score data from linear (conventional) tests in which each testee is administered all items, and items are scored into more than two categories. Three models are available: the graded normal and logistic ogive models (Samejima, 1969), and the nominal logistic model (Bock, 1972). In LINPSCO only maximum likelihood scoring is available, and subscale scoring is not possible.

All three programs compute both test information and response pattern information values when maximum likelihood scoring is used. Response pattern information (Samejima, 1973) provides an estimate of the precision of measurement for a specified response pattern and can be used to compare the quality of trait estimates derived from specific test administration and/or scoring procedures (e.g., Bejar, Weiss, & Gialluca, 1977).

NUMERICAL PROCEDURES

Dichotomous Data

Maximum Likelihood

The numerical procedure for maximum likelihood scoring of dichotomous data consists of two stages. In the first stage an initial estimate is sought by the bisection method. Once this initial estimate is obtained, it is refined further by the Newton-Raphson method.

The bisection routine begins in the interval ±5.00. If the sign of the first derivative of the likelihood function during the first iteration is the same when evaluated at 5.00 and at -5.00, a value of 0.0 is returned as the initial estimate of θ. Otherwise, five additional iterations are performed. After the sixth iteration, the width of the interval has been reduced to $10/(2^6) = 10/64 = .15$. The midpoint of that interval is the initial estimate which is then refined further by Newton-Raphson iterations of the form:

$$\hat{\theta}_{m+1} = \hat{\theta}_m - (f'/f'') \quad , \qquad [3]$$

where

$\hat{\theta}_{m+1}$ is the new estimate,

$\hat{\theta}_m$ is the estimate from the last iteration,

f' is the first derivative of the log-likelihood function evaluated at $\hat{\theta}_m$, and

f'' is the second derivative of the log-likelihood function evaluated at $\hat{\theta}_m$.

This iterative process is continued until $|\hat{\theta}_{m+1} - \hat{\theta}_m| < .005$. If that criterion has not been met at the end of 50 iterations, the case is said to be nonconvergent.

Formulas for derivatives. Let $v = \{u_g, g=1, 2, \ldots n\}$ be a response vector such that

$$u_g = \begin{cases} 1 \text{ if the item is answered correctly} \\ 0 \text{ if the item is answered incorrectly.} \end{cases}$$

Note that for scoring purposes, the response vector does not include rejected or omitted items. The probability that $u_g=1$ for a given value of θ and item parameters a_g, b_g, and c_g is given by

$$P_g(\theta) = c_g + (1-c_g)[1 + e^{-1.7a_g(\theta - b_g)}]^{-1} \qquad [4]$$

for the logistic ogive model and by

$$P_g(\theta) = c_g + (1-c_g)\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta-b_g)} e^{t^2/2}dt$$

$$= c_g + (1-c_g)\,\Phi\,[a_g(\theta-b_g)] \qquad [5]$$

for the normal ogive model, where Φ stands for the standard cumulative normal distribution.

The log-likelihood function for the response vector is

$$L_v(\theta) = \sum_g \log P_g(\theta)^{u_g}\, Q_g(\theta)^{1-u_g}$$

$$= \sum_g [u_g \log P_g(\theta) + (1-u_g)\log Q_g(\theta)] \qquad [6]$$

where $Q_g(\theta) = 1-P_g(\theta)$ and $P_g(\theta)$ is given by either Equation 4 or 5.

In general, the first and second derivatives of the log-likelihood function of a response vector are given by

$$\frac{\partial L_v(\theta)}{\partial \theta} = \sum_g \left\{ u_g \left[\left(\frac{1}{P_g(\theta)} \right) \left(\frac{\partial P_g(\theta)}{\partial \theta} \right) \right] + (1-u_g) \left[\left(\frac{1}{Q_g(\theta)} \right) \left(\frac{\partial Q_g(\theta)}{\partial \theta} \right) \right] \right\} \qquad [7]$$

and

$$\frac{\partial^2 L_v(\theta)}{\partial \theta^2} = \sum_g \left\{ u_g \left[\left(\frac{1}{P_g^2(\theta)} \right) \left(\frac{\partial P_g(\theta)}{\partial \theta} \right)^2 - \left(\frac{1}{P_g(\theta)} \right) \left(\frac{\partial^2 P_g(\theta)}{\partial \theta} \right) \right] \right.$$

$$\left. + (1-u_g) \left[\left(\frac{1}{Q_g^2(\theta)} \right) \left(\frac{\partial Q_g(\theta)}{\partial \theta} \right)^2 + \left(\frac{1}{Q_g(\theta)} \right) \left(\frac{\partial^2 Q_g(\theta)}{\partial \theta^2} \right) \right] \right\} . \qquad [8]$$

For the logistic ogive model, after simplification and letting $x = 1.7a_g(\theta-b_g)$, these expressions are

$$\frac{\partial L_v(\theta)}{\partial \theta} = -1.7 \sum_g \left[\frac{a_g e^x}{1+e^x} \right] + 1.7 \sum_g \left[\frac{u_g a_g e^x}{c_g + e^x} \right] \qquad [9]$$

and

$$\frac{\partial^2 L_v(\theta)}{\partial \theta^2} = -2.89 \sum_g \left[\frac{a_g^2 e^x}{(1+e^x)^2} \right] + 2.89 \sum_g \left[\frac{u_g a_g^2 c_g e^x}{(c_g+e^x)^2} \right] . \qquad [10]$$

For the normal ogive model, letting $x = -a_g^2(\theta-b_g)^2/2$, the corresponding expressions are

$$\frac{L_v(\theta)}{\partial \theta} = \sum_g \left[\frac{u_g (2\pi)^{-1/2} a_g (1-c_g) e^x}{c_g + (1-c_g)\Phi[a_g(\theta-b_g)]} - \frac{(1-u_g)(2\pi)^{-1/2}(1-c_g) a_g e^x}{1-\{c_g+(1-c_g)\Phi[a_g(\theta-b_g)]\}} \right] \qquad [11]$$

and

$$\frac{\partial^2 L_v(\theta)}{\partial\theta^2} = \sum_g \left\{ u_g \left[- \frac{[(2\pi)^{-1/2}(1-c_g)a_g e^x]^2}{\{[c_g-(1-c_g)\Phi[a_g(\theta-b_g)]\}^2} - \frac{(2\pi)^{-1/2}a_g^3(\theta-b_g)(1-c_g)e^x}{\{c_g-(1-c_g)\Phi[a_g(b_g-\theta)]\}} \right] + \right.$$

$$\left. \sum_g (1-u_g) \left[- \frac{[(2\pi)^{-1/2}(1-c_g)a_g e^x]^2}{1-\{c_g+(1-c_g)\Phi[a_g(\theta-b_g)]\}^2} + \frac{(2\pi)^{-1/2}a_g^3(\theta-b_g)(1-c_g)e^x}{1-\{c_g+(1-c_g)\Phi[a_g(\theta-b_g)]\}} \right] \right\} \quad [12]$$

Computation of information. With maximum likelihood scoring, two measures of information are computed for each response pattern. One is response pattern information (Samejima, 1973) denoted by $\hat{I}(\hat{\theta})$; the other is test information (Birnbaum, 1968; Samejima, 1969) denoted by $I(\hat{\theta})$. Test information is defined as the expected value of the second derivative of the log-likelihood function, i.e.,

$$I(\hat{\theta}) = -\sum_g \left[E \left\{ \frac{\partial^2 \log L_v(\theta)}{\partial\theta^2} \right\} \right] . \quad [13]$$

Response pattern information, on the other hand, is defined by

$$\hat{I}(\hat{\theta}) = -\sum_g \left[\frac{\partial^2 \log L_v(\theta)}{\partial\theta^2} \right] , \quad [14]$$

that is, the "observed," as opposed to expected, value of the second derivative of the log-likelihood function evaluated at $\hat{\theta}$.

These two measures of information will be the same for models in which there is a sufficient statistic for the response vector. In particular, this is true in the one- and two-parameter logistic ogive models. It is also true for the "zero" parameter normal ogive model, i.e., when the items are parallel. The value of $\hat{I}(\hat{\theta})$ for a given response pattern is simply the value of the second derivative of the log-likelihood function at the last iteration, i.e., evaluated at the estimated value of θ.

$I(\hat{\theta})$ is computed by

$$I(\hat{\theta}) = \sum_g \frac{\{P_g'(\hat{\theta})\}^2}{P_g(\hat{\theta})\{1.0-P_g(\hat{\theta})\}} , \quad [15]$$

where $P_g(\theta)$ is given, in general, by Equation 4 for the logistic ogive model and by Equation 5 for the normal ogive model.

For the normal ogive model,

$$P'_g(\hat{\theta}) = \frac{a_g(1-c_g)}{\sqrt{2\pi}} \left[e^{-a_g^2[\theta-b_g]^2/2}\right] ; \qquad [16]$$

and for the logistic ogive model,

$$P'_g(\hat{\theta}) = \frac{1.7a_g(1-c_g)e^{1.7a_g(\hat{\theta}-b_g)}}{[1+e^{1.7a_g(\hat{\theta}-b_g)}]^2} . \qquad [17]$$

Standard error. The standard error of measurement associated with $\hat{\theta}$ is computed as $1/\sqrt{\hat{I}(\hat{\theta})}$, that is, the reciprocal square root of response pattern information evaluated at $\hat{\theta}$.

Bayesian Scoring

The Bayesian scoring procedure used by LINDSCO and ADADSCO is derived from Owen's (1975) sequential adaptive testing strategy. However, since the present application assumes that the test items have already been administered, only the scoring aspect is of interest.

The procedure makes the assumption that the prior distribution of θ is normal, with mean μ_o=0.0 and variance σ_o^2=1.00, where subscript $_o$ denotes the fact that no items have yet been administered. After the m^{th} item is administered, the mean and variance of the posterior density function are computed according to the following equations. If the response to the $m+1^{th}$ item is correct,

$$\mu_{m+1} = E(\theta|1) = \mu_m + (1-c_g)\left(\frac{\sigma^2_m}{\sqrt{\frac{1}{a^2_g} + \sigma^2_m}}\right)\left(\frac{\phi(D)}{c_g+(1-c_g)\Phi(-D)}\right) \qquad [18]$$

and

$$\sigma^2_{m+1} = \mathrm{var}(\theta|1) = \sigma^2_m \left\{ 1-\left(\frac{1-c_g}{1+\frac{1}{a^2_g\sigma^2_m}}\right)\left(\frac{\phi(D)}{A}\right)\left(\frac{(1-c_g)\phi(D)}{A} - D\right)\right\} . \qquad [19]$$

Following an incorrect answer,

$$\mu_{m+1} = E(\theta|0) = \mu_m - \left(\sqrt{\frac{\sigma^2_m}{\frac{1}{a^2_g} + \sigma^2_m}}\right)\left(\frac{\phi(D)}{\Phi(D)}\right) \qquad [20]$$

and

$$\sigma^2_{m+1} = \text{var}(\theta|0) = \sigma^2_m \left\{1 - \left(\frac{\phi(D)}{1 + \frac{1}{a^2_g \sigma^2_m}}\right)\left(\frac{\frac{\phi(D)}{\Phi(D)} + D}{\Phi(D)}\right)\right\}. \qquad [21]$$

In Equations 18 through 21 (from Owen, 1975),

$\phi(D)$ is the normal probability density function,
$\Phi(D)$ is the cumulative normal distribution function,

$$D = \frac{b_g - \mu_m}{\sqrt{\frac{1}{a^2_g} + \sigma^2_m}} \quad \text{, and} \qquad [22]$$

$$A = c_g + (1-c_g)\ \Phi\ (-D)\,. \qquad [23]$$

After the last item has been administered, the posterior mean is the estimated θ and the posterior variance is a measure of the error associated with that estimate. Because the posterior distribution after every item is administered is approximated by a normal distribution in this procedure, there is a certain amount of inaccuracy in the estimate. Moreover, the resulting scores are order dependent (Sympson, 1977), i.e., if a response vector were to be scored after rearranging the items, the resulting θ estimate would be slightly different.

Computation of Expected Proportion of Correct Answers

The expected proportion of correct answers (EXPTOT) is defined as

$$\text{EXPTOT} = \Sigma_g P_g(\hat{\theta})/NI\ , \qquad [24]$$

where $P_g(\hat{\theta})$ is computed from Equation 4 for the logistic ogive model and Equation 5 for normal ICCs. *NI* is the number of items on which the estimate of θ is based. EXPTOT is simply an estimate of the true score associated with $\hat{\theta}$ (Lord & Novick, 1968, p. 387).

Polychotomous Data

LINPSCO is capable of scoring polychotomous data when item parameters have been estimated according to a graded model of either normal or logistic ogive form (Samejima, 1969) or according to Bock's (1972) nominal logistic model. For the graded model, the numerical procedure consists of a bisection stage of six iterations followed by Newton-Raphson iterations. For the nominal logistic model, the initial estimate obtained from the bisection stage is refined further by the secant method rather than by Newton-Raphson iterations.

In each case, the bisection phase begins in the interval ±5.00. During the first iteration, if the sign of the first derivative of the log-likelihood function is the same when evaluated at 5.00 and at -5.00, a value of 0.0 is returned as the initial estimate. Otherwise, five additional iterations are performed. After six iterations, the width of the interval is reduced to $10/(2^6)=10/64=.15$. The midpoint of that interval is taken as the initial estimate.

The Newton-Raphson procedure used with the graded models refines the initial estimate with iterations of the form shown in Equation 3. This iterative procedure is continued until $|\hat{\theta}_{m+1} - \hat{\theta}_m|$ is less than .005 or the number of iterations is greater than 50. The secant procedure is similar to Newton-Raphson iterations, except that f'' in Equation 3 is an approximation to the second derivative of the log-likelihood function given by

$$f'' = \frac{f'(\hat{\theta}_m) - f'(\hat{\theta}_{m-1})}{(\hat{\theta}_m - \hat{\theta}_{m-1})} \quad . \qquad [25]$$

Graded Models

Let $v = \{x_g,\ g=1,\ 2,\ \ldots\ \}$ be a response vector exclusive of omitted and rejected items such that

$$x_g = \begin{cases} 1 = \text{if the "best response was given} \\ 2 = \text{if the second "best" response was given} \\ \cdot \\ \cdot \\ \cdot \\ m_g\text{-}1 \text{ if the next to worst response was given} \\ m_g \text{ if the worst response was given.} \end{cases}$$

For the graded logistic ogive model, the probability that x_g takes one of the values between 1 and m_g is given in general by

$$P_{x_g}(\theta) = P_{x_g} = [1 + e^{y x_g}]^{-1} - [1 + e^{y x_g - 1}]^{-1} \quad , \qquad [26]$$

where

$$y_{x_g} = -a_g D(\theta - b_{x_g}) \quad , \tag{27}$$

$$y_{x_g-1} = -a_g D(\theta - b_{x_g-1}) \quad , \text{ and} \tag{28}$$

$D = 1.7$ is a scaling factor.

When $x_g = 1$,

$$P_{x_g} = [1 + e^{y_{x_g}}]^{-1} \quad . \tag{29}$$

When $x_g = m_g$,

$$P_{x_g} = 1 - [1 + e^{y_{x_g-1}}]^{-1} . \tag{30}$$

For the graded normal ogive model, the probability that x_g takes one of the values between 1 and m_g is given in general by

$$P_{x_g}(\theta) = P_{x_g} = (2\pi)^{-\frac{1}{2}} \int_{y_{x_g-1}}^{y_{x_g}} e^{t^2/2} dt$$

$$= \Phi\left[y_{x_g}\right] - \Phi\left[y_{x_g-1}\right] \quad , \tag{31}$$

where

$$y_{x_g} = a_g(\theta - b_{x_g}) \tag{32}$$

and

$$y_{x_g-1} = a_g(\theta - b_{x_g-1}) \quad . \tag{33}$$

When $x_g = 1$,

$$P_{x_g} = (2\pi)^{-1} \int_{-\infty}^{y_{x_g}} e^{t^2/2} dt \quad . \tag{34}$$

When $x_g = m_g$,

$$P_{x_g} = 1 - (2\pi)^{-1} \int_{-\infty}^{y_{x_g}-1} e^{t^2/2} dt \quad . \qquad [35]$$

The log-likelihood function for a given response vector is given by

$$L_v(\theta) = \log \prod_g P_{x_g}^{r_{x_g}}$$

$$= \sum_g r_{x_g} [\log P_{x_g}] \ , \qquad [36]$$

where

$$r_{x_g} = \begin{cases} 1 \text{ if the } x_g^{\text{th}} \text{ response category is chosen} \\ 0 \text{ otherwise} \end{cases} .$$

The general first derivative of the log-likelihood function is

$$\frac{\partial L_v(\theta)}{\partial \theta} = \sum_g \sum_{x_g} r_{x_g} L_{x_g} \quad . \qquad [37]$$

Samejima (1969) referes to L_{x_g} as the basic function. Since $L_{x_g} = (\partial P_{x_g}/\partial\theta)(P_{x_g})^{-1}$

$$\frac{\partial L_v(\theta)}{\partial \theta} = \sum_g \sum_{x_g} r_{x_g} \frac{\partial P_{x_g}/\partial\theta}{P_{x_g}} \quad . \qquad [38]$$

The general second derivative of the log-likelihood function is given by

$$\frac{\partial^2 L_v(\theta)}{\partial \theta^2} = \sum_g \sum_{x_g} r_{x_g} [-(L_{x_g})^2 + \frac{\partial^2 P_{x_g}/\partial\theta^2}{P_{x_g}}] \quad . \qquad [39]$$

Specifically, for the graded logistic ogive model,

$$\frac{\partial L_v(\theta)}{\partial \theta} = \sum_g \sum_{x_g} a_g 1.7\{1 - P^*_{x_g} - P^*_{x_g-1}\} \qquad [40]$$

and

$$\frac{\partial^2 L_v(\theta)}{\partial \theta^2} = \sum_g \sum_{x_g} 2.89 a_g^2 \frac{r_{x_g}}{P_{x_g}} \left[-\left\{ (L_{x_g})^2 + (2Q^*_{x_g} - 1)(P^*_{x_g} Q^*_{x_g}) - \right.\right.$$

$$\left.\left. (2Q^*_{x_g - 1} - 1)(P^*_{x_g - 1} Q^*_{x_g - 1}) \right\} \right] \quad , \qquad [41]$$

where

$$P^*_{x_g} = [1 + e^{-1.7 a_g (\theta - b_{x_g})}]^{-1} \quad , \qquad [42]$$

$$P^*_{x_g - 1} = [1 + e^{-1.7 a_g (\theta - b_{x_g - 1})}]^{-1} \quad , \qquad [43]$$

$$P^*_o = 0 \quad , \qquad [44]$$

$$P^*_{m_g} = 1 \quad , \qquad [45]$$

$$Q^*_{x_g} = 1 - P^*_{x_g} \quad , \text{ and} \qquad [46]$$

$$Q^*_{x_g - 1} = 1 - P_{x_g - 1} \quad . \qquad [47]$$

For the graded normal ogive model, letting $z_{x_g} = -[a_g^2(\theta - b_{x_g})^2] / 2$, the corresponding expressions are

$$\frac{\partial L_v(\theta)}{\partial \theta} = \sum_g \left[\sum_{x_g} \left(\frac{r_{x_g} a_g}{\sqrt{2\pi}} [e^{z_{x_g}} - e^{z_{x_g - 1}}] \right) / P_{x_g}(\theta) \right] \qquad [48]$$

The second derivative is given by

$$\frac{\partial^2 L_v(\theta)}{\partial \theta^2} = \sum_g \left\{ \sum_{x_g} r_{x_g} \left[- (L_{x_g})^2 \right] + \right.$$

$$\left. \left[\frac{-a_g^3}{\sqrt{2\pi}} \{ (\theta - b_{x_g}) e^{z_{x_g}} - (\theta - b_{x_g - 1}) e^{z_{x_g - 1}} \} \right] / P_{x_g}(\theta) \right\} \qquad [49]$$

When $x_g=1$, $e^{z_{x_g}-1}=0$, and P_{x_g} is given by Equation 34; when $x_g=m_g$, $e^{z_{x_g}}=0$, and P_{x_g} is given by Equation 35.

Nominal Logistic Model

For the nominal logistic model, the probability of x_g, given θ, is given by

$$P_{x_g}(\theta) = P_{x_g} = e^{(\alpha_{x_g}\theta+\beta_{x_g})} \Big/ \sum_{s=1}^{m_g} e^{(\alpha_s\theta+\beta_s)} \quad , \qquad [50]$$

where α_s and β_s are the slope and intercept parameter for the s^{th} response category.

The secant method requires only the first derivative of the log-likelihood function. That derivative is

$$\frac{\partial L_v(\theta)}{\partial\theta} = \sum_g \frac{\sum_{s=1}^{m_g} r_{x_g}(\alpha_{x_g}-\alpha_s)e^{\alpha_s\theta+\beta_s}}{\sum_{s=1}^{m_g} e^{(\alpha_s\theta+\beta_s)}} \quad . \qquad [51]$$

Computation of Information

Response pattern information is computed as the value of the second derivative at the last iteration. For the nominal logistic model, that value is an approximation. Test information is computed from the general formula given by Samejima (1969),

$$I(\hat{\theta}) = \sum_g \sum_{x_g} (\partial P_{x_g}/\partial\theta)^2 P_{x_g} \quad . \qquad [52]$$

This expression involves only the first derivative of the response model. The appropriate expressions are listed below.

For the graded normal ogive model,

$$\frac{\partial P_{x_g}}{\partial\theta} = \frac{a_g}{\sqrt{2\pi}}\left[e^{z_{x_g}} - e^{z_{x_g}-1}\right] \quad , \qquad [53]$$

where $z_{x_g} = -[a_g^2(\theta-b_{x_g})^2]/2$.

For the graded logistic ogive model,

$$\frac{\partial P_{x_g}}{\partial \theta} = 1.7\, a_g [P^*_{x_g}(1-P^*_{x_g}) - P^*_{x_g-1}(1-P^*_{x_g-1})] , \qquad [54]$$

where $P^*_{x_g} = [1+e^{-1.7a(\theta-b_{x_g})}]^{-1}$.

For the nominal logistic model,

$$\frac{\partial P_{x_g}}{\partial \theta} = \frac{[e^{(\alpha_{x_g}\theta + \beta_{x_g})} \sum_{s=1}^{m_g} e^{(\alpha_{s_g}\theta + \beta_{s_g})} (\alpha_{x_g} - \alpha_s)]}{\sum_{s=1}^{m_g} e^{(\alpha_{s_g}\theta + \beta_{s_g})^2}} . \qquad [55]$$

USE OF THE PROGRAMS

Input

For each of the programs, three types of input are required:

1. The *Program Parameters*, which consist of specifications as to the number of items in the pool, the options chosen, the scoring key, and so forth.
2. The *Item Pool*, which contains the item parameter estimates on as many as 600 items for LINDSCO and ADADSCO, and 100 items for LINPSCO.
3. The *Test Response Data* consists of testee name and identification number and each testee's item responses. For LINDSCO, item responses need not be dichotomized beforehand; for ADADSCO, they must be dichotomized unless a key is provided as part of the item pool. For ADADSCO, the number of items attempted and the identification number of each item attempted must also be supplied as part of the test response data. For LINPSCO, the test response data must be supplied in such a way that the first category corresponds to the "best" response, while the last category corresponds to the "worst" response, based on previously obtained item parameterization data.

Testee response data containing all correct or incorrect answers cannot be scored by maximum likelihood. If such a response pattern is found, a message is printed, and the estimated θ is set to 10.00 if all responses are correct and to -10.00 if all responses are incorrect. The information is set to 0.0 in both cases. Response patterns with all answers correct or incorrect present no problem for Bayesian scoring, and they are processed normally; however, a message is still printed. Appendix B gives examples of the use of each of these programs.

Table 6

Input Program Parameters for LINDSCO, ADADSCO,and LINPSCO: Card Set 1

Columns	LINDSCO	ADADSCO	LINPSCO
1-4 (I4)	INUP, number of items in item pool. 600 is the maximum.	INUP, same as LINDSCO	Number of items in the pool. Maximum is 100.
5-8 (I4)	M, number of items in test. 300 is the maximum	MMAX, maximum number of items administered. 60 is the maximum.	M, number of items in the test. Maximum is 50.
9	blank	blank	blank
10 (I1)	OPT1 1 = Punch the item parameter estimates corresponding to the items in the test.	OPT1 1 = Print the item parameter estimates corresponding to the items administered (this is done only for the first 10 testees).	OPT1 1 = Punch the item parameter estimates corresponding to the items in the test.
11 (I1)	OPT2 1 = the item pool consists of M items, i.e., there will be no searching of items in the pool.	not used	OPT2 1 = the item pool consists of M items, i.e., there will be no searching of items in the pool.
12 (I1)	OPT3 If 1, 2, or 3 item parameters will be edited; see "Editing of item parameters."	OPT3, same as LINDSCO	not used
13 (I1)	OPT4, scoring algorithms and response model: 1 = maximum likelihood normal ogive 2 = maximum likelihood logistic 3 = Owen's Bayesian normal ogive	OPT4, same as LINDSCO	OPT4, response model: 1 = graded logistic ogive 2 = graded normal ogive 3 - nominal logistic

14-18 (F5.2)	TS, for Bayesian scoring. This is the prior mean of θ. Not used in maximum likelihood scoring.	TS, same as LINDSCO	D, scaling parameter for graded logistic model. If blank, will be set by default to 1.0; otherwise will usually be set by the user to 1.7.
19-23 (F5.2)	TSS, for Bayesian scoring. This is the prior standard deviation of θ. Not used in maximum likelihood scoring.	TSS, same as LINDSCO	not used
24-28 (F5.2)	AMAX, value of the a parameter. Used in editing. See "Editing of item parameters."	AMAX, same as LINDSCO	not used
29-33 (F5.2)	BMIN, lowest value of the b parameter. Used in editing parameter estimates. See "Editing of item parameters."	BMIN, same as LINDSCO	not used
34-38 (F5.2)	BMAX, highest value of the b parameter. Used in editing parameter estimates. See "Editing of item parameters."	BMAX, same as LINDSCO	not used
39-43 (F5.2)	CMAX, value of the c parameter. Used in editing parameter estimates. See "Editing of item parameters."	CMAX, same as LINDSCO	not used
44-45 (I2)	blank	IFLAG, code for correct response	
46-47 (I2)	IOMIT, code for omitted response.	IOMIT, same as LINDSCO	IOMIT, same as LINDSCO
48-80	blank	blank	blank

Program Parameters

Table 6 describes the input program parameters for all three programs, using Card Set 1 (all numeric information is right justified). After Card Set 1, the program parameter and input for each of the three programs differs, as indicated below.

LINDSCO (Card Set 2-10).

Card Set 2 (8A10).	The variable format for the item pool is punched on this card, using I-fields (see Item Pool below).
Card Set 3 (16I5).	Punch in five-column fields the item identification number of the items in the test in the same order in which they appear in the test. Continue on as many cards as necessary.
Card Set 4 (80I1).	A "1" in a given column is punched to omit a specified item from all computations, e.g., if the 10th item is to be omitted, punch "1" in column 10; if the 100th item is to be omitted, punch "1" in column 20 of the second card. Continue on as many cards as necessary.
Card Set 5 (80I1).	This card contains the scoring key for the test. In general, the n^{th} column contains the key for the n^{th} item, as in Card Set 4. Continue on as many cards as necessary.
Card Set 6 (8A10).	Variable format for reading the subject information and test response data (see Test Response Data below for field type specifications).
Card Set 7 (8A10).	The description of the run is written on three cards. The three cards must be included even if they are blank.
Card 8 (I5).	Punch the number of subscales to be scored in columns 1-5; maximum is 25. If no subscales are to be scored, punch "0" in column 5; in that case, this is the last card set.
Card Set 9 (2I5). (Omit if the number of subscales is 0.)	For each scale, punch the following information: Columns 1-5: Number of items in subscale (maximum is 60). Columns 6-10: Scale number. Repeat for each subscale beginning on a new card.
Card Set 10 (16I5). (Omit if the number of subscales is 0.)	Punch in five-column fields the item identification number of the items in the subscales. Continue on as many cards as necessary. Repeat for each subscale, beginning on a new card for each subscale.

ADADSCO (Card Set 2-5).

Card Set 2 (8A10). Variable format for item pool, using I-fields. It must be contained on one card (see Item Pool below).

Card Set 3 (16I5). Columns 1-5: The number of items to be omitted, i.e., excluded from the computations. If none, punch "0" in column 5.
Columns 6-10 and subsequent five-column fields: The item identification numbers of items to be omitted. Continue on as many cards as necessary. If more than one card is necessary, begin punching on the second card in columns 1-5.

Card Set 4 (8A10). Variable input format for reading subject information and test response data. It must be contained on one card (see Test Response Data below for field type specifications).

Card Set 5 (8A10). Description of the run is written on three cards. These cards are required, even if they are left blank.

LINPSCO (Card Set 2-7).

Card Set 2 (8A10). Variable format for the item pool. It must be contained on one card (see Item Pool below for field type specifications).

Card Set 3 (80I1). Punch in the n^{th} column the number of response categories minus 1 for the n^{th} item. Continue on as many cards as necessary.

Card Set 4 (16I5). Punch in five-column fields the item identification numbers of the items in the test. The numbers must appear in the same order as the items appear in the test. Continue on as many cards as necessary.

Card Set 5 (80I1). The information on this card is used to omit specified items from the computations. To omit the n^{th} item, punch a "1" in the n^{th} column of this card; otherwise, punch "0." If no items are to be omitted, punch as many zeros as there are items in the test. Continue on as many cards as necessary.

Card Set 6 (8A10). Variable format for subject information and test response data. It must be contained on one card (see Test Response Data below for field type specifications).

Card Set 7 (8A10). Description of the run is written on three cards.

Item Pool

LINDSCO and ADADSCO. To score the response data, a file containing the item pool item parameter estimates must be prepared beforehand and placed in a file called IPOOL. The file consists of a line for each item in the pool with the following information:

1. A *unique* item number;
2. Estimate of the a parameter;
3. Estimate of the b parameter;
4. Estimate of the c parameter;
5. Correct alternative for this item, i.e., the keyed response.

For LINDSCO, only Items 1 through 4 must be supplied; for ADADSCO, Item 5 must be supplied also, although it could be a "dummy" key (e.g., a blank), since the data may already be scored (see columns 44-45 for Card Set 1).

The exact format of this information is not critical, since it is read with a user-specified variable format. However, the following limitations must be observed: (1) the information must be read in the above order; (2) the item number must be read in integer mode; (3) the item parameter estimates must be read in floating point; and (4) the key, if ADADSCO is being used, must be read in integer mode.

A typical format for LINDSCO could be

```
(10X,I4,3F10.2)  .
```

For ADADSCO, a typical format might be

```
(10X,I4,3F10.2,I2)  .
```

All three parameter estimates must be read even if the user is using a one- or two-parameter model. This presents no difficulties, however, since in the case of, say, a two-parameter model, the third parameter is 0 for all items. This may be accomplished by reading blanks or zeros, or by editing item parameter estimates (see below).

The number of items in the pool may range from the number of items in the test, M, to 600. If the item pool for LINDSCO consists of only the items being scored in the test, then OPT2 should be set to 1. This indicates to the program that items do not have to be searched. On the other hand, if the pool consists of items in addition to those used in the present test, then OPT2 should be set to 0. This instructs the program to search for the item and to retrieve the corresponding item parameters. For both LINDSCO and ADADSCO, if at least one of the items being called for is not found in the pool, the program prints a message; and the unavailable item is treated as an omitted item.

Editing of item parameter estimates. LINDSCO and ADADSCO have several options to edit item parameter estimates. If OPT3=1, the program checks that the item parameter estimates are within certain bounds. For the discrimination (a) parameter, the program checks to see if the estimate exceeds AMAX; if it does, it is set to AMAX. For the difficulty (b) parameter, if the estimate is below BMIN, it is set to BMIN; if it is above BMAX, it is set to BMAX. For the "guessing" (c) parameter, the program checks to see if the estimate exceeds CMAX; if it does, it is set to CMAX. If the user wants to edit only one or two parameters, the limits of the other parameters should be chosen so that the editing has no effect.

A more radical form of editing is also possible. If OPT3=2, then in addition to the editing caused by OPT3=1, the program sets all c parameter estimates to CMAX. If CMAX=0.0, this implies that a two-parameter model is in effect. If OPT3=3, then in addition to the editing caused by OPT3=1 and OPT3=2, the program sets all a parameter estimates to AMAX.

LINPSCO. For polychotomous scoring, the item pool consists of the following information for graded normal and logistic ogive models:

1. A *unique* item identification number;
2. The "discrimination" parameter, which is common to all response categories;
3. m_g-1 "difficulty" parameters, where m_g is the number of response categories in the g^{th} item. Since m_g can be at most 10, there would be at most 9 difficulty parameters.

The exact format for reading this information is not crucial, since it is read by a user-supplied format statement. However, the following restrictions must be observed: (1) the identification number is read first, in integer mode; (2) next, the estimated discrimination parameter is read in floating point mode; (3) the m_g-1 "difficulty" parameters are read next, with the difficulty of the best alternative followed by the second best alternative, and so forth.

Since the program allows the number of categories to differ from item to item, the format should be specified so that it can read the information for the item with the most response categories. For example, if in a given test, the maximum number of response categories is seven, then there should be at most six difficulty parameters. The format for such pools might be as follows:

```
(I4,6X,F5.2/10X,6F5.2)  .
```

In this format the item identification number is read in the I4 field; the discrimination parameter is read next in format F5.2; and the six difficulty values are read from the next card, beginning in column 11.

For the nominal logistic model, the item information is read in the following order:

1. A *unique* item identification number;
2. m_g "slope" parameters; and
3. m_g intercept parameters.

Differing from the graded models, in the nominal model there is a pair of parameters (a slope and an intercept) associated with *each* response category. Since the response categories are not ordered in the nominal model, the order in which the parameters are read is unimportant. However, the ordinal position in which the parameters appear in the pool must correspond with the integer associated with that response category. As in the graded models, the format should be able to read the information for the item with most response categories. For example, if the maximum number of response categories is five, the format

could be

(I4,16X,5F5.2,5X,5F5.2) .

In this format, the item identification number is read in the I4 field; next, the five slope parameters are read in 5F5.2; and finally, the five intercept parameters are read in the last set of 5F5.2 fields.

Test Response Data

Data for all testees must be on a file called DATA. The structure of this file differs slightly for each of the programs. In all cases, however, the last record of DATA must be an end-of-record marker.

LINDSCO. This program requires that for each individual the following information be provided on DATA:

1. Name,
2. Identification number, and
3. Responses to the test items.

The exact format of this information is not critical, since it is read with a user-supplied variable format; but the information must appear in the above order. Two words are used for testee name; thus, name should be read with two alphanumeric words, e.g., 2A10. This allows for up to 20 characters. The testee identification is read with an alphanumeric field of at least 1 column, e.g., A1, A9. Test item responses are read with an integer format, e.g., 20I1.

The test data may be raw item responses (i.e., the number of the alternatives chosen) or scored (i.e., 0 for incorrect and 1 for correct). However, in either case, a scoring key must be provided (see Card Set 5 for LINDSCO). The key will contain the number of the correct alternative if raw data are read. If the data are already scored, a "dummy" key full of "1's" must be provided.

Omitted items are indicated by the integer IOMIT (see columns 46 and 47 of Card Set 1 for LINDSCO). For raw data, this will normally be an integer greater than the number of alternatives. Similarly, for scored (0-1) data IOMIT must be an integer greater than 1.

ADADSCO. The program requires that the following information be provided on DATA for each individual:

1. Name,
2. Identification number,
3. Number of items answered by the testee (i.e., number of items attempted),
4. Item identification numbers of items attempted, and
5. Responses to the test items.

This information is read in the above order with a user-supplied variable format; thus, the exact format is not critical. However, the following limitations must be observed. Even though the number of items administered usually varies across individuals in an adaptive test, this program assumes that the data record for each testee is formally the same (i.e., that there is the same number of data

lines per testee and that these lines contain similar information). Thus, if the maximum number of items taken by anyone is MMAX (see Card Set 1 for ADADSCO), but any particular testee takes *M* items, where *M*<MMAX, then that testee's record should be "padded" to MMAX items. This can be accomplished by leaving an appropriate number of blank fields. The name is dimensioned for two words so the format should allow for two words, e.g., 2A10. The identification number is read with an alphanumeric format, e.g., A8. The number of items is read in integer mode. The item identification numbers and item responses are also read in integer mode. Note that in reading the item identification numbers and the item responses, the format should read MMAX of each, even if some of these will be blank for a given individual.

As an example assume that MMAX was 25; then the variable format could be
(2A10,A10,I2/20I4/5I4/25I1).
In this format, the name, testee identification, and *M* are read from the first card; the item identification numbers are read from the next two cards; and finally, the item responses are read from the fourth card. Note that for testees attempting 20 items or less, the third card will be blank.

The item responses may be scored or raw data. For scored data, the responses have been reduced to three categories: correct, incorrect, and omitted. In this case, IFLAG should be set to the integer corresponding to the correct code, and IOMIT should be set to the code for omitted responses. Note that if IFLAG>0, the program ignores the key read as part of the item pool. For raw data, the key will have been read as part of the item pool; IFLAG must therefore be set to 0. IOMIT will still be operational, however; and it must be set to an integer other than the highest numbered response alternative.

LINPSCO. The DATA file is similar to LINDSCO's with the exception that the item responses must include the response category chosen by the testee for a given item. For graded models, the convention that the best response category be coded "1," second best "2," and so forth, must be obeyed. For the nominal logistic model, this convention does not apply; but care must be taken so that a category's response code matches the ordinal position of that category in the IPOOL file. For either graded or nominal data, the code for omitted responses should be an integer greater than the maximum number of response categories.

Output

Four kinds of output are produced by each program: program parameters, item parameters, computational messages, and testee data.

Program Parameters

The output consists of the information in Card Set 1, the description of the run, and the variable formats for reading the item pool and the testee's raw data.

Item Parameters

LINDSCO and LINPSCO. The output consists of item identification number, scoring key, rejection key (i.e., whether or not the item was included in the computations), and the item parameter estimates. If the estimates have been

edited, the edited values will be printed. An option (see column 10 of Card Set 1) permits all of this information to be punched as well. If subscale scoring has been requested, the item identification number of the items in each subscale will be printed.

ADADSCO. The user has the option, but only for the first 10 testees, to print the following: testee's name and identification number; and for each item attempted, the item identification number, the response to that item, and the item parameter estimates.

Computational Messages

The program will print a testee's name and identification number if (1) a response pattern is found with all items correct or incorrect, excluding omitted or rejected items; (2) a zero score has been obtained; or (3) it was not possible to achieve convergence in scoring the testee's responses. For polychotomous data, a perfect or zero vector occurs if the testee responds with the best or worst response categories in all attempted items, exclusive of omitted or rejected items. If an item is not found in the pool or has extreme parameter estimates, an informative message is printed.

The number of testees read and the number of convergence failures are also printed. If Bayesian scoring has been requested, the number of nonconvergent cases will be zero.

Testee Data

LINDSCO. For each testee, the following information is written on a file called TAPE3:

1. Name;
2. Testee identification number;
3. Scale number, or in the case of total score, a "T";
4. Proportion of items answered correctly;
5. Maximum likelihood or Bayesian estimate of θ;
6. The response pattern information for maximum likelihood scoring or the posterior variance of θ for Bayesian scoring;
7. The number of items used in the estimation of θ, excluding items rejected, omitted, or not found;
8. The test information associated with the estimated θ (for Bayesian scoring, the information is computed using the normal ogive model);
9. The true score corresponding to the estimated θ;
10. For maximum likelihood scoring,
    a. The number of Newton-Raphson iterations needed to achieve convergence and
    b. The standard error of θ.

The format used for writing this information for total scores is (X,2A10,A9,*T*,F5.2,2F7.2,I4,2F7.2,I4,F7.2) .

The subscale results are written with (X,2A10,A9,I2,F5.2,2F7.2,I4,2F7.2,I4,F7.2) .

ADADSCO. The same information is written as that for LINDSCO with the exception of the scale number. The format is
(X,2A10,A9,F5.2,2F7.2,I4,2F7.2,I4,F7.2) .

LINPSCO. For LINPSCO, the following information is written:

1. Name;
2. Testee identification number;
3. Proportion of "best" responses;
4. Maximum likelihood estimate of θ;
5. The response pattern information;
6. The number of items used in the estimation of θ excluding items rejected, omitted, or not found;
7. The number of iterations needed to achieve convergence.
8. The test information associated with the estimated θ;
9. Estimated standard error of measurement.

AVAILABILITY

FORTRAN source code listings of the three programs are in Appendix C (LINDSCO), Appendix D (ADADSCO), and Appendix E (LINPSCO). Copies of the FORTRAN source code are available on cards or tape at nominal cost from

Psychometric Methods Program
Department of Psychology
University of Minnesota
75 East River Road
Minneapolis, Minnesota 55455

Telephone: 612-376-7378

Potential users of these programs should note that the programs were written for Control Data Corporation CYBER series computers. Because of the large word size of the CYBER computers, accurate computation on other computers may require the use of double-precision arithmetic. Minimal additional modifications required may include (1) modification of A10 fields to smaller sizes used by other computers and (2) modification of FORTRAN statements unique to the CYBER series computers.

REFERENCES

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A047495)

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

Bock, R. D. Estimating latent parameter and ability when responses are scored on two or more nominal categories. Psychometrika, 1972, 37, 29-51.

Kendall, M. G., & Stuart, A. The advanced theory of statistics (Vol 2). New York: Hafner, 1961.

Kolakowski, D., & Bock, R. D. A FORTRAN IV program for maximum likelihood item analysis and test scoring: Normal ogive model (Research Memo No. 12). Chicago: University of Chicago, Department of Education, Statistics Laboratory, 1970.

Kolakowski, D., & Bock, R. D. LOGOG: Maximum likelihood item analysis and test scoring: Logistic model for multiple responses (Research Memo No. 13). Chicago: University of Chicago, Department of Education, Statistics Laboratory, 1972.

Lawley, D. N. On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 1943, 61, 273-287.

Lord, F. M. Some test theory for tailored testing. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper & Row, 1970.

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

Owen, R. J. A Bayesian sequential procedure for sequential response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika, Monograph Supplement No. 17, 1969.

Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. Psychometrika, 1973, 28, 221-234.

Sympson, J. B. Estimation of latent trait status in adaptive testing procedures. In D. J. Weiss (Ed.), Applications of computerized adaptive testing (Research Report 77-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977. (NTIS No. AD A038114)

Weiss, D. J. (Ed.) Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975. (NTIS No. AD A018675)

Weiss, D. J. Computerized ability testing, 1972-1975 (Final Report). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1976. (NTIS No. A024516).

Wood, R. L., Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (Research Memo 76-6). Princeton, NJ: Educational Testing Service, 1976.

Wright, B. D., & Mean, R. J. CALFIT: Sample free item calibration with a Rasch measurement model (Research Memo No. 18). Chicago: University of Chicago, Department of Education, Statistical Laboratory, 1976.

Appendix A

Item Parameter Estimation Programs

Program Name	Model	Reference
LOGIST	Three-parameter logistic ogive	Wood, R. L., Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (Research Memorandum 76-6). Princeton, NJ: Educational Testing Service, 1976.
NORMOG	Three-parameter normal ogive	Kolakowski, D., Bock, R. A. A FORTRAN IV program for maximum likelihood item analysis and test scoring: Normal ogive model (Research Memo No. 12). Chicago: University of Chicago, Department of Education, Statistics Laboratory, 1970.
ESTEM	Three-parameter logistic or normal ogive	No program documentation available. For a description of the procedure see Urry, V. W., Ancillary estimators for the parameters of mental test models. Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center, 1974.
BICAL	One-parameter logistic ogive	Wright, B. D., & Mead, R. J. CALFIT: Sample-free item calibration with a Rasch measurement model (Research Memorandum No. 18). Chicago: University of Chicago, Department of Education, Statistical Laboratory, 1976.
LOGOG	Graded normal and logistic ogive, nominal logistic	Kolakowski, D., & Bock, R. D. LOGOG: A FORTRAN IV program for maximum likelihood item analysis and test scoring: Logistic model for multiple response (Research Memorandum No. 13). Chicago: University of Chicago, Department of Education, Statistical Laboratory, 1972.

Appendix B:

Examples of Program Use

The following examples serve to illustrate the use of each of the three programs. These results should also be useful in testing the accuracy of the results of the programs in different installations.

LINDSCO

The IPOOL file for these examples was

1	1.000	-2.000	.25
2	1.500	-1.000	.25
3	1.000	0.000	.25
4	1.500	1.000	.25
5	1.000	2.000	.25

The DATA file is also shown below. The first field contains the names; the second, the subject identification; and the third, the response patterns.

Name	I.D.	Responses
A00	1	00000
A01	2	00001
A02	3	00010
A03	4	00011
A04	5	00100
A05	6	00101
A06	7	00110
A07	8	00111
A08	9	01000
A09	10	01001
A10	11	01010
A11	12	01011
A12	13	01100
A13	14	01101
A14	15	01110
A15	16	01111
A16	17	10000
A17	18	10001
A18	19	10010
A19	20	10011
A20	21	10100
A21	22	10101
A22	23	10110
A23	24	10111
A24	25	11000
A25	26	11001
A26	27	11010
A27	28	11011
A28	29	11100
A29	30	11101
A30	31	11110
A31	32	11111

Example 1. This example illustrates the use of the normal ogive model (OPT4=1) for a five-item test (m=5) with a pool containing five items (INUP=5). The example also illustrates the use of parameter editing (OPT3=1) in which BMAX=BMIN=0.0, which in effect sets all b parameter estimates to 0.0. AMAX=2.00, which means that if there were a parameter estimates greater than 2, they would be set to 2.00. CMAX=.10, which means that any c parameter estimates greater than .10 will be edited to .10. This example also illustrates the use of subscales. The program parameter cards for this example were

```
   5    5  111 0.00 1.00 2.00 0.00 0.00  .10  4444
(9X,I1,3F10.3)
    1     2     3     4     5
00000
11111
(2A4,A2,10X,5I1)
RUNS BASED ON ALL POSSIBLE RESPONSE VECTORS FOR A FIVE ITEM TEST


    1
    3     1
    2     3     5
```

The output corresponding to this example is shown on the following pages.

* *

LINDSCO

LINEAR DICHOTOMUS SCORING WITH THREE PARAMETER MODELS

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MPLS. MINN. 55455

```
INUP    =    5
MMAX    =    5
IOMIT   = 4444
OPT1    =   -0
OPT2    =    1
OPT3    =    1
OPT4    =    1
TS      =    0
TSS     = 1.00
AMAX    = 2.00
BMAX    =    0
BMIN    =    0
CMAX    =  .10
VARIABLE FORMAT FOR POOL=(9X,I1,3F10.3)
VARIABLE FORMAT FOR DATA=(2A4,A2,10X,5I1)
RUNS BASED ON ALL POSSIBLE RESPONSE VECTORS FOR A FIVE ITEM TEST
```

ITEMS IN SUBSCALE NO= 1

2 3 5

* *

* *

ITEM ID S	KEYS	REJECTIONS	A	B	C
1	1	0	1.00	0	.10
2	1	0	1.50	0	.10
3	1	0	1.00	0	.10
4	1	0	1.50	0	.10
5	1	0	1.00	0	.10

* *

```
SUBJECT =A00     ID = 1     HAS NO  ANSWERS CORRECT IN TOTAL SCALE
SUBJECT =A00     ID = 1     HAS NO  ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A02     ID = 3     HAS NO  ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A13     ID =14     HAS ALL ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A15     ID =16     HAS ALL ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A16     ID =17     HAS NO  ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A18     ID =19     HAS NO  ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A29     ID =30     HAS ALL ANSWERS CORRECT IN SUBSCALE      1
SUBJECT =A31     ID =32     HAS ALL ANSWERS CORRECT IN TOTAL SCALE
SUBJECT =A31     ID =32     HAS ALL ANSWERS CORRECT IN SUBSCALE      1

CASES  READ=   32 CASES NOT  CONVERGED=     0
```

Contents of TAPE3:

Testee Name	Testee Identification Number	Subscale: T=Total; 1=Scale 1	Proportion Correct	Maximum Likelihood Estimate of θ	Response Pattern Information	Number of Items	Test Information Associated with $\hat{\theta}$	Expected Proportion Correct	Number of Iterations	Estimated Standard Error of Measurement
A00	1	T	0	-10.00	0	5	0	0	0	
A00	1	1	0	-10.00	0	5	0	0	0	
A01	2	T	.20	-1.06	1.48	5	1.07	.20	2	.82
A01	2	1	.33	-.65	1.51	3	1.36	.30	3	.81
A02	3	T	.20	-.95	.58	5	1.36	.22	5	1.31
A02	3	1	0	-10.00	0	5	0	0	0	
A03	4	T	.40	-.31	3.35	5	3.40	.42	3	.55
A03	4	1	.33	-.65	1.51	3	1.36	.30	3	.81
A04	5	T	.20	-1.06	1.48	5	1.07	.20	2	.82
A04	5	1	.33	-.65	1.51	3	1.36	.30	3	.81
A05	6	T	.40	-.50	3.08	5	2.84	.35	2	.57
A05	6	1	.67	.05	2.33	3	2.23	.57	2	.66
A06	7	T	.40	-.31	3.35	5	3.40	.42	3	.55
A06	7	1	.33	-.65	1.51	3	1.36	.30	3	.81
A07	8	T	.60	.07	3.98	5	3.94	.58	2	.50
A07	8	1	.67	.05	2.33	3	2.23	.57	2	.66
A08	9	T	.20	-.95	.58	5	1.36	.22	5	1.31
A08	9	1	.33	-.30	1.83	3	1.95	.43	3	.74
A09	10	T	.40	-.31	3.35	5	3.40	.42	3	.55
A09	10	1	.67	.40	2.03	3	2.11	.71	2	.70
A10	11	T	.40	-.12	3.63	5	3.78	.50	2	.52
A10	11	1	.33	-.30	1.83	3	1.95	.43	3	.74
A11	12	T	.60	.27	3.71	5	3.88	.67	2	.52
A11	12	1	.67	.40	2.03	3	2.11	.71	2	.70
A12	13	T	.40	-.31	3.35	5	3.40	.42	3	.55
A12	13	1	.67	.40	2.03	3	2.11	.71	2	.70
A13	14	T	.60	.07	3.98	5	3.94	.58	2	.50
A13	14	1	1.00	10.00	0	5	0	0	0	
A14	15	T	.60	.27	3.71	5	3.88	.67	2	.52
A14	15	1	.67	.40	2.03	3	2.11	.71	2	.70
A15	16	T	.80	.75	2.82	5	2.94	.83	2	.60
A15	16	1	1.00	10.00	0	5	0	0	0	
A16	17	T	.20	-1.06	1.48	5	1.07	.20	2	.82
A16	17	1	0	-10.00	0	5	0	0	0	
A17	18	T	.40	-.50	3.08	5	2.84	.35	2	.57
A17	18	1	.33	-.65	1.51	3	1.36	.30	3	.81
A18	19	T	.40	-.31	3.35	5	3.40	.42	3	.55
A18	19	1	0	-10.00	0	5	0	0	0	
A19	20	T	.60	.07	3.98	5	3.94	.58	2	.50
A19	20	1	.33	-.65	1.51	3	1.36	.30	3	.81
A20	21	T	.40	-.50	3.08	5	2.84	.35	2	.57
A20	21	1	.33	-.65	1.51	3	1.36	.30	3	.81
A21	22	T	.60	-.11	3.94	5	3.79	.50	2	.50
A21	22	1	.67	.05	2.33	3	2.23	.57	2	.66
A22	23	T	.60	.07	3.98	5	3.94	.58	2	.50
A22	23	1	.33	-.65	1.51	3	1.36	.30	3	.81
A23	24	T	.80	.46	3.82	5	3.62	.74	2	.51
A23	24	1	.67	.05	2.33	3	2.23	.57	2	.66
A24	25	T	.40	-.31	3.35	5	3.40	.42	3	.55
A24	25	1	.33	-.30	1.83	3	1.95	.43	3	.74
A25	26	T	.60	.07	3.98	5	3.94	.58	2	.50
A25	26	1	.67	.40	2.03	3	2.11	.71	2	.70
A26	27	T	.60	.27	3.71	5	3.88	.67	2	.52
A26	27	1	.33	-.30	1.83	3	1.95	.43	3	.74
A27	28	T	.80	.75	2.82	5	2.94	.83	2	.60
A27	28	1	.67	.40	2.03	3	2.11	.71	2	.70
A28	29	T	.60	.07	3.98	5	3.94	.58	2	.50
A28	29	1	.67	.40	2.03	3	2.11	.71	2	.70
A29	30	T	.80	.46	3.82	5	3.62	.74	2	.51
A29	30	1	1.00	10.00	0	5	0	0	0	
A30	31	T	.80	.75	2.82	5	2.94	.83	2	.60
A30	31	1	.67	.40	2.03	3	2.11	.71	2	.70
A31	32	T	1.00	10.00	0	5	0	0	0	
A31	32	1	1.00	10.00	0	5	0	0	0	

Example 2. This example is identical to Example 1 except that the Bayesian scoring routine was used (OPT4=3) instead of the maximum likelihood normal ogive. Only the scoring results are shown.

Testee Name	Testee Identification Number	Subscale: T=Total; 1=Scale 1	Proportion correct	Bayesian estimate of θ	Bayesian Posterior Variance	Number of Items	Test Information Associated with θ̂	Expected Proportion Correct		
A00	1	T	0	-1.23	.31	5	.71	.17	0	0
A00	1	1	0	-1.04	.39	3	.69	.21	0	0
A01	2	T	.20	-.88	.35	5	1.58	.24	0	0
A01	2	1	.33	-.55	.44	3	1.55	.34	0	0
A02	3	T	.20	-.83	.38	5	1.73	.25	0	0
A02	3	1	0	-1.04	.39	3	.69	.21	0	0
A03	4	T	.40	-.34	.41	5	3.33	.41	0	0
A03	4	1	.33	-.55	.44	3	1.55	.34	0	0
A04	5	T	.20	-.89	.30	5	1.56	.23	0	0
A04	5	1	.33	-.55	.41	3	1.54	.33	0	0
A05	6	T	.40	-.51	.32	5	2.83	.35	0	0
A05	6	1	.67	.02	.44	3	2.22	.56	0	0
A06	7	T	.40	-.37	.33	5	3.23	.40	0	0
A06	7	1	.33	-.55	.41	3	1.54	.33	0	0
A07	8	T	.60	.11	.35	5	3.95	.60	0	0
A07	8	1	.67	.02	.44	3	2.22	.56	0	0
A08	9	T	.20	-.78	.28	5	1.92	.26	0	0
A08	9	1	.33	-.36	.39	3	1.87	.41	0	0
A09	10	T	.40	-.40	.30	5	3.14	.38	0	0
A09	10	1	.67	.23	.43	3	2.22	.65	0	0
A10	11	T	.40	-.25	.30	5	3.53	.44	0	0
A10	11	1	.33	-.36	.39	3	1.87	.41	0	0
A11	12	T	.60	.20	.33	5	3.93	.64	0	0
A11	12	1	.67	.23	.43	3	2.22	.65	0	0
A12	13	T	.40	-.40	.27	5	3.15	.38	0	0
A12	13	1	.67	.22	.42	3	2.22	.64	0	0
A13	14	T	.60	-.02	.28	5	3.89	.54	0	0
A13	14	1	1.00	.98	.50	3	1.35	.88	0	0
A14	15	T	.60	.19	.30	5	3.93	.63	0	0
A14	15	1	.67	.22	.42	3	2.22	.64	0	0
A15	16	T	.80	.69	.33	5	3.10	.81	0	0
A15	16	1	1.00	.96	.50	3	1.35	.88	0	0
A16	17	T	.20	-.82	.25	5	1.76	.25	0	0
A16	17	1	0	-1.04	.39	3	.69	.21	0	0
A17	18	T	.40	-.50	.26	5	2.84	.35	0	0
A17	18	1	.33	-.55	.44	3	1.55	.34	0	0
A18	19	T	.40	-.38	.27	5	3.19	.39	0	0
A18	19	1	0	-1.04	.39	3	.69	.21	0	0
A19	20	T	.60	.01	.29	5	3.91	.55	0	0
A19	20	1	.33	-.55	.44	3	1.55	.34	0	0
A20	21	T	.40	-.50	.24	5	2.84	.35	0	0
A20	21	1	.33	-.55	.41	3	1.54	.33	0	0
A21	22	T	.60	-.17	.25	5	3.69	.48	0	0
A21	22	1	.67	.02	.44	3	2.22	.56	0	0
A22	23	T	.60	-.01	.26	5	3.90	.55	0	0
A22	23	1	.33	-.55	.41	3	1.54	.33	0	0
A23	24	T	.80	.39	.27	5	3.73	.71	0	0
A23	24	1	.67	.02	.44	3	2.22	.56	0	0
A24	25	T	.40	-.37	.24	5	3.23	.40	0	0
A24	25	1	.33	-.36	.39	3	1.87	.41	0	0
A25	26	T	.60	-.02	.25	5	3.89	.54	0	0
A25	26	1	.67	.23	.43	3	2.22	.65	0	0
A26	27	T	.60	.17	.27	5	3.94	.62	0	0
A26	27	1	.33	-.36	.39	3	1.87	.41	0	0
A27	28	T	.80	.61	.30	5	3.29	.79	0	0
A27	28	1	.67	.23	.43	3	2.22	.65	0	0
A28	29	T	.60	-.03	.24	5	3.88	.54	0	0
A28	29	1	.67	.22	.42	3	2.22	.64	0	0
A29	30	T	.80	.34	.26	5	3.80	.69	0	0
A29	30	1	1.00	.98	.50	3	1.35	.88	0	0
A30	31	T	.80	.59	.32	5	3.33	.78	0	0
A30	31	1	.67	.22	.42	3	2.22	.64	0	0
A31	32	T	1.00	1.20	.37	5	1.72	.92	0	0
A31	32	1	1.00	.98	.50	3	1.35	.88	0	0

Example 3. This example illustrates the use of the maximum likelihood logistic scoring routine (e.g., OPT4=2) without subscale scoring. Only the scoring results are shown.

Testee Name	Testee Identification Number	Subscale: T=Total; 1=Subscale 1	Proportion Correct	Maximum Likelihood Estimate of θ	Response Pattern Information	Number of Items	Test Information Associated with $\hat{\theta}$	Expected Proportion Correct	Number of Iterations	Estimated Standard Error of Measurement
AU0	1	T	0	-10.00	0	5	0	0	0	
AU1	2	T	.20	-1.05	1.20	5	.90	.20	2	.91
AU2	3	T	.20	-99.99	-99.99	5	-99.99	-99.99	99	-99.99
AU3	4	T	.40	-.29	3.59	5	3.68	.42	3	.53
AU4	5	T	.20	-1.05	1.20	5	.90	.20	2	.91
AU5	6	T	.40	-.46	3.32	5	2.95	.35	1	.55
AU6	7	T	.40	-.29	3.59	5	3.68	.42	3	.53
AU7	8	T	.60	.07	4.49	5	4.46	.58	2	.47
AU8	9	T	.20	-99.99	-99.99	5	-99.99	-99.99	99	-99.99
AU9	10	T	.40	-.29	3.59	5	3.68	.42	3	.53
A10	11	T	.40	-.11	4.01	5	4.26	.50	2	.50
A11	12	T	.60	.25	4.22	5	4.26	.66	2	.49
A12	13	T	.40	-.29	3.59	5	3.68	.42	3	.53
A13	14	T	.60	.07	4.49	5	4.46	.58	2	.47
A14	15	T	.60	.25	4.22	5	4.26	.66	2	.49
A15	16	T	.80	.71	2.74	5	2.72	.83	2	.60
A16	17	T	.20	-1.05	1.20	5	.90	.20	2	.91
A17	18	T	.40	-.46	3.32	5	2.95	.35	1	.55
A18	19	T	.40	-.29	3.59	5	3.68	.42	3	.53
A19	20	T	.60	.07	4.49	5	4.46	.58	2	.47
A20	21	T	.40	-.46	3.32	5	2.95	.35	1	.55
A21	22	T	.60	-.10	4.53	5	4.28	.50	2	.47
A22	23	T	.60	.07	4.49	5	4.46	.58	2	.47
A23	24	T	.80	.45	3.70	5	3.70	.74	2	.52
A24	25	T	.40	-.29	3.59	5	3.68	.42	3	.53
A25	26	T	.60	.07	4.49	5	4.46	.58	2	.47
A26	27	T	.60	.25	4.22	5	4.26	.66	2	.49
A27	28	T	.80	.71	2.74	5	2.72	.83	2	.60
A28	29	T	.60	.07	4.49	5	4.46	.58	2	.47
A29	30	T	.80	.45	3.70	5	3.70	.74	2	.52
A30	31	T	.80	.71	2.74	5	2.72	.83	2	.60
A31	32	T	1.00	10.00	0	5	0	0	0	

ADADSCO

IPOOL for this example consisted of 10 items:

1	1.00	-2.00	.25 1
2	1.25	-1.50	.25 1
3	1.50	-1.00	.25 1
4	1.75	-0.50	.25 1
5	1.00	0.00	.25 1
6	1.25	0.50	.25 1
7	1.50	1.00	.25 1
8	1.75	1.50	.25 1
9	1.00	2.00	.25 1
10	1.25	2.50	.25 1

The data for the 16 subjects used in the example are shown below:

```
1  A    10
1011101110
 1 2 3 4 5 6 7 8 910
2  B     4
1010
 2 4 6 8
3  C     5
11111
 1 2 3 4 5
4  D     5
00000
 6 7 8 910
5  E     8
11000000
 1 2 3 5 6 7 8 9
6  F     2
10
 3 7
7  G     6
011111
 1 2 4 6 8 9
8  H     9
101010101
 1 2 3 4 5 6 7 8 9
9  I     7
1100110
 1 3 4 5 6 810
10 J     3
110
 4 8 9
```

```
11 K     8
10110110
 2 3 4 6 7 8 910
12 L     9
101111111
 1 2 3 4 5 7 8 910
13 M     5
01001
 1 3 5 7 9
14 N     6
000001
 4 5 6 7 8 9
15 P     7
1011110
 1 2 4 5 6 7 8
16 Q     6
011011
 2 4 5 7 810
```

The program control cards for this example were

```
  10   10 1012 0.00 1.00 2.00 0.00 0.00  .10 1 3
(8X,I2,3F10.2,I2)
    0
(A2,1X,2A2,I2,/10I1,/10I2)
DESCRIPTION
```

In this example the maximum number of items attempted by anyone was 10 (MMAX=10). Although the code for omitted items was 3 (IOMIT=3), IFLAG=1, which means the key to each item was read from IPOOL; however, in this case it was 1 for all items. OPT1=1 means that item information for the first 10 subjects will be printed. Editing of item parameters was requested (OPT3=1). The scoring algorithm was maximum likelihood logistic.

The entire output for this example is shown on the following pages.

ADAOSCO

ADAPTIVE DICHOTOMUS SCORING WITH THREE PARAMETER MODELS

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MPLS. MINN. 55455

```
INUP   =   10
MMAX   =   10
IOMIT  =    3
IFLAG  =    1
OPT1   =    1
OPT2   =    0
OPT3   =    1
OPT4   =    2
IS     =    0
TSS    = 1.00
AMAX   = 2.00
BMAX   =    0
BMIN   =    0
CMAX   =  .10
VARIABLE FORMAT FOR POOL=(8X,I2,3F10.2,I2)
VARIABLE FORMAT FOR DATA=(A2,1X,2A2,I2,/10I1,/10I2)
DESCRIPTION
```

Item Identification Number	Scored Answer	Discrimination Parameter (a)	Difficulty Parameter (b)	"Guessing" Parameter (c)
	1 A			
1	1	1.00	0	.10
2	0	1.25	0	.10
3	1	1.50	0	.10
4	1	1.75	0	.10
5	1	1.00	0	.10
6	0	1.25	0	.10
7	1	1.50	0	.10
8	1	1.75	0	.10
9	1	1.00	0	.10
10	0	1.25	0	.10
	2 B			
2	1	1.25	0	.10
4	0	1.75	0	.10
6	1	1.25	0	.10
8	0	1.75	0	.10
	3 C			
1	1	1.00	0	.10
2	1	1.25	0	.10
3	1	1.50	0	.10
4	1	1.75	0	.10
5	1	1.00	0	.10
	SUBJECT C		ID= 3	HAS ALL ANSWERS RIGHT
	4 D			
6	0	1.25	0	.10
7	0	1.50	0	.10
8	0	1.75	0	.10
9	0	1.00	0	.10
10	0	1.25	0	.10

```
      SUBJECT D                          ID= 4        HAS NO RIGHT ANSWERS
          5       E
  1   1          1.00          0         .10
  2   1          1.25          0         .10
  3   0          1.50          0         .10
  5   0          1.00          0         .10
  6   0          1.25          0         .10
  7   0          1.50          0         .10
  8   0          1.75          0         .10
  9   0          1.00          0         .10
          6       F
  3   1          1.50          0         .10
  7   0          1.50          0         .10
          7       G
  1   0          1.00          0         .10
  2   1          1.25          0         .10
  4   1          1.75          0         .10
  6   1          1.25          0         .10
  8   1          1.75          0         .10
  9   1          1.00          0         .10
          8       H
  1   1          1.00          0         .10
  2   0          1.25          0         .10
  3   1          1.50          0         .10
  4   0          1.75          0         .10
  5   1          1.00          0         .10
  6   0          1.25          0         .10
  7   1          1.50          0         .10
  8   0          1.75          0         .10
  9   1          1.00          0         .10
          9       I
  1   1          1.00          0         .10
  3   1          1.50          0         .10
  4   0          1.75          0         .10
  5   0          1.00          0         .10
  6   1          1.25          0         .10
  8   1          1.75          0         .10
 10   0          1.25          0         .10
         10       J
  4   1          1.75          0         .10
  8   1          1.75          0         .10
  9   0          1.00          0         .10
      CASES  READ=    16 CASES NOT  CONVERGED=     0
```

Testee Name	Testee Identification Number	Proportion Correct	Maximum Likelihood Estimate of θ	Response Pattern Information	Number of Items	Test Information Associated with $\hat{\theta}$	Expected Proportion Correct	Number of Iterations	Estimated Standard Error of Measurement
A	1	.70	.34	9.65	10	9.65	.71	3	.32
B	2	.50	-.22	5.06	4	4.66	.43	2	.44
C	3	1.00	10.00	0	5	0	0	0	0
D	4	0	-10.00	0	5	0	0	0	0
E	5	.25	-.78	3.15	8	2.69	.24	1	.56
F	6	.50	-.09	2.57	2	2.57	.50	2	.62
G	7	.83	.79	3.13	6	3.12	.87	2	.57
H	8	.56	-.09	9.85	9	9.55	.50	2	.32
I	9	.57	.05	8.03	7	8.03	.58	2	.35
J	10	.67	.43	3.18	3	3.20	.77	2	.56
K	11	.63	.13	9.62	8	9.62	.62	2	.32
L	12	.89	.89	4.02	9	4.02	.89	3	.50
M	13	.40	-.29	3.59	5	3.68	.42	3	.53
N	14	.17	-1.08	1.40	6	.85	.18	2	.84
P	15	.71	.25	7.55	7	7.52	.68	2	.36
Q	16	.67	.23	7.03	6	7.03	.67	2	.38

LINPSCO

Following are sample runs from LINPSCO using graded models and the nominal logistic model.

Graded models. The IPOOL file for these examples was

```
1 1.5  3.0  2.0  4.5  3.0
2 1.5  2.0  1.0  3.0  1.5
3 1.5  1.0  0.0  1.5  0.0
4 1.5  0.0 -1.0  0.0 -1.5
5 1.5 -1.0 -2.0 -1.5 -3.0
6 1.5 -2.0 -3.0 -3.0 -4.5
```

The DATA file, including subject identification and item responses, was as follows. Note that in coding the item responses, a "1" indicated the "best" response and a "3" indicated the "poorest," as specified by the item difficulty parameters in IPOOL.

```
 1        333211
 2        332111
 3        321212
 4        112121
 5        112233
 6        222222
 7        122221
 8        322223
 9        111333
10        111222
11        333222
12        331111
13        222111
14        211111
```

The following is an example of the logistic graded model (OPT4=1) with a 1.7 scaling factor. In this example the *b* parameters for the items were taken from columns 3-4 of the IPOOL file. The option and format cards for this example were

```
   6    0201 1 1.70                                    4
(I1,1X,F3.1,2(1X,F4.1))
222222
   1    2    3    4    5    6
000000
(2A7,A1,6I1)
EXAMPLE RUN OF THE LOGISTIC GRADED MODEL--USES THE FIRST PAIR OF B
PARAMETERS FROM ITEM POOL
```

The output from this run was as follows:

```
                                    LINPSCO


        LINEAR POLYCHOTOMUS SCORING WITH TWO PARAMETER MODELS


                                PSYCHOMETRICS METHODS PROGRAM
                                DEPARTMENT OF PSYCHOLOGY
                                UNIVERSITY OF MINNESOTA
                                MPLS.  MINN. 55455


        INUP   =   6
        MMAX   =   6
        IOMIT  =   4
        OPT1   =   0
        OPT2   =   1
        OPT4   =   1
        MAXCAT =   2
        D      = 1.7
        VARIABLE FORMAT FOR POOL =(I1,1X,F3.1,2(1X,F4.1))
        VARIABLE FORMAT FOR DATA =(2A7,A1,6I1)
        EXAMPLE RUN OF THE LOGISTIC GRADED MODEL--USES THE FIRST PAIR OF B
        PARAMETERS FROM ITEM POOL

ITEM ID =      1      REJECTION =  0
  A:     1.50
  B:     3.00  2.00

ITEM ID =      2      REJECTION =  0
  A:     1.50
  B:     2.00  1.00

ITEM ID =      3      REJECTION =  0
  A:     1.50
  B:     1.00     0

ITEM ID =      4      REJECTION =  0
  A:     1.50
  B:        0 -1.00

ITEM ID =      5      REJECTION =  0
  A:     1.50
  B:    -1.00 -2.00

ITEM ID =      6      REJECTION =  0
  A:     1.50
  B:    -2.00 -3.00



CASES READ =   14  CASES NOT CONVERGED =     0
```

Testee Identification	Proportion of "Best" Responses (Coded 1)	Maximum Likelihood Estimate of θ	Response Pattern Information	Number of Items Used to Estimate θ	Number of Iterations	Test Information Associated with $\hat{\theta}$	Estimated Standard Error of Measurement
1	.33	-.50	4.72	6	2	4.25	.46
2	.50	.50	4.72	6	2	4.25	.46
3	.33	.35	3.69	6	3	4.28	.52
4	.67	1.52	2.63	6	3	4.11	.62
5	.33	-.00	4.21	6	3	4.41	.49
6	0	-.00	5.16	6	2	4.41	.44
7	.33	.51	4.88	6	2	4.25	.45
8	0	-.51	4.88	6	2	4.25	.45
9	.50	-.00	.96	6	3	4.41	1.02
10	.50	.52	2.65	6	2	4.25	.61
11	0	-1.51	4.86	6	2	4.11	.45
12	.50	-.00	4.20	6	3	4.41	.49
13	.50	1.51	4.86	6	2	4.11	.45
14	.83	2.69	3.12	6	2	2.67	.57

Following is an example of use of the normal ogive graded model (OPT4=2) using the same DATA and IPOOL as the previous example. In this example the b parameters for the items were taken from columns 5 and 6 of the IPOOL file. Input control cards for this example were

```
   6    0201 2                                          4
(I1,1X,F3.1,11X,F4.1,1X,F4.1)
222222
    1    2    3    4    5    6
000000
(2A7,A1,6I1)
EXAMPLE RUN OF THE NORMAL OGIVE GRADED MODEL--USES SECOND PAIR OF B
PARAMETERS FROM ITEM POOL
```

Output was as follows:

LINPSCO

LINEAR POLYCHOTOMUS SCORING WITH TWO PARAMETER MODELS

PSYCHOMETRICS METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MPLS. MINN. 55455

```
INUP   =   6
MMAX   =   6
IOMIT  =   4
OPT1   =   0
OPT2   =   1
OPT4   =   2
MAXCAT =   2
VARIABLE FORMAT FOR POOL =(I1,1X,F3.1,11X,F4.1,1X,F4.1)
VARIABLE FORMAT FOR DATA=(2A7,A1,6I1)
EXAMPLE RUN OF THE NORMAL OGIVE GRADED MODEL--USES SECOND PAIR OF B
PARAMETERS FROM ITEM POOL

ITEM ID =      1      REJECTION =  0
  A:     1.50
  B:     4.50  3.00

ITEM ID =      2      REJECTION =  0
  A:     1.50
  B:     3.00  1.50

ITEM ID =      3      REJECTION =  0
  A:     1.50
  B:     1.50     0

ITEM ID =      4      REJECTION =  0
  A:     1.50
  B:        0 -1.50

ITEM ID =      5      REJECTION =  0
  A:     1.50
  B:    -1.50 -3.00

ITEM ID =      6      REJECTION =  0
  A:     1.50
  B:    -3.00 -4.50


CASES READ =   14  CASES NOT CONVERGED =     0
```

1	.33	-.75	2.97	6	2	3.36	.58
2	.50	.75	2.97	6	2	3.36	.58
3	.33	-.04	8.00	6	2	3.42	.35
4	.67	1.85	8.05	6	2	3.39	.35
5	.33	-.00	11.98	6	2	3.42	.29
6	0	-.00	11.58	6	2	3.42	.29
7	.33	1.09	9.50	6	2	3.39	.32
8	0	-1.09	9.50	6	2	3.39	.32
9	.50	-.00	12.80	6	1	3.42	.28
10	.50	.75	12.29	6	2	3.36	.29
11	0	-2.25	5.18	6	2	3.35	.44
12	.50	-.00	3.19	6	2	3.42	.56
13	.50	2.25	5.18	6	2	3.35	.44
14	.33	3.92	2.05	6	2	2.25	.70

Nominal logistic model. Following is an example of use of the nominal logistic model (OPT4=3). The DATA file was

```
10001 38 9946454134111122442211121114111411411112111111111
10002 381038345411144114141411441111141111141114111414111
10003 38 940005424141121114412141211141111441111111411241
10005 381042450424222444424211142442421222421212121222241
10004 38 995434441414444414144444141114441111141141141144411
10005 381045285411144151214551121122111131111155555555551
10004 381055445443511111514111144111511141141111111114111
10008 38 993349423245514323314352415555555555555555555555555
10009 381053444414212412232134413124214124221114321222554
10010 381051509444444111444511141111141141411111111111144141
10011 381052945411111111511411111111111111111111111111111111
10012 381038492442323133433311422225131325112214451124215 1
10013 381054044441444414411144144411144141144111424151111 1
10015 38 952455414451213254441141111111121241134411112141
10014 381059835421124114214414444121121412114111154411
10015 381051295421114111244111141111111111111111111111141
10014 381059528524413211314211142512111111411114111111 31451
10018 381052484411441114132411141411424443144111411411 5141
10019 381052825441441114424111243311344241124112131111411 4
10020 251041143444444414144111444444414144441411141441444444
```

IPOOL was

3417	0.000000	0.000000	0.000000	0.000000
	0.000000	0.000000	0.000000	0.000000
3242	0.870169	0.165539	-1.212113	0.176405
	0.843205	0.266821	-2.116617	1.006591
3251	1.039257	0.007821	-1.157598	0.110520
	0.906771	-0.711543	-1.432918	1.237689
3204	0.876227	0.313705	-1.132723	-0.059209
	0.510097	-0.183194	-1.493435	1.166532
3422	0.849335	-0.056859	-0.936221	0.143745
	0.947914	-0.510533	-1.591918	1.154558
3411	0.997044	-0.057076	-1.046348	0.106380
	0.223949	-0.511724	-0.842978	1.130754
3421	0.903633	0.452463	-1.315750	-0.040285
	1.324394	0.196635	-2.289410	0.768330
3277	0.935577	0.125247	-1.013213	-0.047610
	1.754989	-0.277937	-1.882968	0.405916
3408	1.267512	0.165917	-1.188968	-0.244462
	0.722900	-0.503622	-1.456565	1.237267
3230	0.406782	0.236527	-0.675647	0.032338
	-0.165035	-0.081050	-0.762658	0.968744
3012	0.914806	-0.087146	-0.807633	-0.020027
	0.880125	-0.361430	-1.133896	0.615221
3010	0.000000	0.000000	0.000000	0.000000
	0.000000	0.000000	0.000000	0.000000
3405	1.213009	-0.160534	-1.049275	-0.005201
	0.703199	-0.472368	-1.271001	1.040191
3213	0.941547	0.120564	-1.216924	0.154792
	1.616184	-0.227161	-2.405661	1.014638
3207	1.196701	0.289311	-1.692915	0.234903
	2.322618	-0.227163	-3.302494	1.207041
3079	1.382173	-0.138215	-1.203993	-0.039965
	1.753215	-0.061868	-2.059620	0.968273
3002	1.160829	-0.009762	-1.500254	0.349208
	0.157054	-0.506044	-1.638385	1.987376
3039	1.234116	-0.075624	-0.934679	-0.223813
	1.111268	0.012128	-1.527785	0.404369
3279	1.023419	0.034131	-0.950948	-0.106622
	0.553385	-0.287433	-1.247386	0.981456
3262	1.603121	-0.068531	-1.228285	-0.308305
	2.697426	-0.693601	-2.469919	0.466094
3220	1.391045	-0.042953	-1.013781	-0.334310
	1.683738	-0.088790	-1.734153	0.139205
3208	1.022427	0.183646	-1.046231	-0.159844
	1.574865	-0.443161	-1.656987	0.525223
3404	0.994791	-0.067707	-1.062704	0.135620
	1.010467	-0.657834	-1.145258	0.786626
3430	1.242329	0.078533	-1.435152	0.114271
	2.364325	-0.659730	-2.638202	1.133657
3006	1.121352	0.164464	-0.962733	-0.343020
	1.620360	-0.262232	-1.632686	0.274608
3237	1.356141	0.216193	-1.205470	-0.366864
	2.263497	-0.343737	-2.311491	0.359781
3022	0.603197	0.002163	-0.812300	0.007000
	0.973607	-0.472746	-1.024659	0.523197
3021	2.042249	0.393252	-1.696127	-0.737373
	3.702481	0.132937	-3.787033	-0.054436
3221	1.377861	0.131896	-1.134625	-0.373132
	2.123626	0.281772	-2.499535	0.094757
3017	1.543139	0.286432	-2.081177	0.249606
	2.042910	0.001334	-3.331305	1.267061
3246	1.564642	-0.007945	-1.298770	-0.257927
	2.293293	0.021343	-2.644136	0.326495
3076	1.283963	0.227021	-0.889192	-0.623795
	3.019197	-0.447594	-2.181694	-0.339909
3215	1.054357	-0.059429	-0.669305	-0.295623
	1.629671	-0.30951	-1.288305	-0.033315
3038	1.453187	0.264191	-1.620217	-0.102162
	1.771329	-0.348563	-2.527377	1.104548
3226	1.119323	-0.063934	-0.823243	-0.292126
	1.279163	-0.303697	-1.057345	0.081860
3239	1.330099	0.036376	-1.773188	0.406713
	2.393186	-0.114603	-3.584869	1.306283
3023	1.466976	-0.073618	-0.823643	-0.589715
	2.233098	-0.331496	-1.535212	-0.371390
3236	1.202460	-0.157735	-0.924427	-0.120298
	1.774003	-0.242566	-1.844767	0.313329
3028	1.066538	-0.131232	-0.860250	-0.069056
	1.633196	-0.843935	-1.242548	0.453307
3024	1.326054	0.287821	-1.859475	0.243501
	1.664725	0.560963	-3.220217	0.994529
3077	1.600690	-0.074231	-1.264068	-0.268341
	2.077034	-0.645261	-2.336619	0.994846
3029	0.691561	0.155307	-1.158516	0.311618
	1.310654	-1.378577	-1.585535	1.653456
3078	1.091246	0.129610	-1.359914	-0.460943
	3.629455	-0.593270	-3.131633	0.095448

Input control cards were

```
  43  22300 3                                    5
(I4,4(1X,F9.6),/,4X,4(1X,F9.6))
3333333333333333333333333333333333333333333
 3417 3251 3422 3421 3277 3408 3010 3405 3213 3079 3002 3220 3404 3430 3021 3221
 3076 3236 3029 3078 3239 3023
10000010000000000000
(2A7,A1,43I1)
EXAMPLE RUN OF THE NOMINAL LOGISTIC MODEL
```

And the output from this run was

LINPSCO

LINEAR POLYCHOTOMUS SCORING WITH TWO PARAMETER MODELS

PSYCHOMETRICS METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MPLS. MINN. 55455

```
INUP   =  43
MMAX   =  22
IOMIT  =   5
OPT1   =   0
OPT2   =   0
OPT4   =   3
MAXCAT =   3
VARIABLE FORMAT FOR POOL =(I4,4(1X,F9.6),/,4X,4(1X,F9.6))
VARIABLE FORMAT FOR DATA=(2A7,A1,43I1)
EXAMPLE RUN OF THE NOMINAL LOGISTIC MODEL
```

```
ITEM ID =      0      REJECTION =  1
  A:       0       0       0       0
  B:       0       0       0       0

ITEM ID =   3251      REJECTION =  0
  A:    1.04    .01 -1.16    .11
  B:     .91   -.71 -1.43   1.24

ITEM ID =   3422      REJECTION =  0
  A:     .85   -.06  -.94    .14
  B:     .95   -.51 -1.59   1.15

ITEM ID =   3421      REJECTION =  0
  A:     .90    .45 -1.32   -.04
  B:    1.32    .20 -2.29    .77

ITEM ID =   3277      REJECTION =  0
  A:     .94    .13 -1.01   -.05
  B:    1.75   -.28 -1.80    .41

ITEM ID =   3408      REJECTION =  0
  A:    1.27    .17 -1.19   -.24
  B:     .72   -.50 -1.46   1.24

ITEM ID =      0      REJECTION =  1
  A:       0       0       0       0
  B:       0       0       0       0

ITEM ID =   3405      REJECTION =  0
  A:    1.22   -.16 -1.05   -.01
  B:     .70   -.47 -1.27   1.04

ITEM ID =   3213      REJECTION =  0
  A:     .94    .12 -1.22    .15
  B:    1.62   -.23 -2.41   1.01

ITEM ID =   3079      REJECTION =  0
  A:    1.38   -.14 -1.20   -.04
  B:    1.75   -.66 -2.06    .97

ITEM ID =   3002      REJECTION =  0
  A:    1.16   -.01 -1.50    .35
  B:     .16   -.51 -1.64   1.99

ITEM ID =   3220      REJECTION =  0
  A:    1.39   -.04 -1.01   -.33
  B:    1.68   -.09 -1.73    .14

ITEM ID =   3404      REJECTION =  0
  A:     .99   -.07 -1.06    .14
  B:    1.02   -.66 -1.15    .79

ITEM ID =   3450      REJECTION =  0
  A:    1.24    .08 -1.44    .11
  B:    2.36   -.86 -2.64   1.13

ITEM ID =   3021      REJECTION =  0
  A:    2.04    .39 -1.70   -.74
  B:    3.71    .13 -3.79   -.05

ITEM ID =   3221      REJECTION =  0
  A:    1.38    .13 -1.13   -.38
  B:    2.12    .28 -2.50    .09

ITEM ID =   3076      REJECTION =  0
  A:    1.29    .23  -.89   -.62
  B:    3.02   -.45 -2.18   -.39

ITEM ID =   3256      REJECTION =  0
  A:    1.20   -.16  -.92   -.12
  B:    1.77   -.24 -1.84    .31

ITEM ID =   3029      REJECTION =  0
  A:     .69    .16 -1.16    .31
  B:    1.31  -1.38 -1.59   1.65

ITEM ID =   3078      REJECTION =  0
  A:    1.69    .13 -1.36   -.46
  B:    3.63   -.59 -3.13    .10

ITEM ID =   3259      REJECTION =  0
  A:    1.33    .04 -1.77    .41
  B:    2.39   -.11 -3.58   1.31

ITEM ID =   3023      REJECTION =  0
  A:    1.49   -.07  -.82   -.59
  B:    2.24   -.33 -1.54   -.37

CASES READ =    20  CASES NOT CONVERGED =      0
```

Testee Identification		Proportion of "Best" Responses	Maximum Likelihood Estimate of θ	Response Pattern Information	Number of Items Used to Estimate θ	Number of Iterations	Test Information Associated with $\hat{\theta}$	Estimated Standard Error of Measurement
8 99484	5	.50	-.63	9.16	20	3	9.12	.33
8103834	5	.65	-.02	6.24	20	3	6.21	.40
8 94000	5	.55	-.33	7.94	20	2	7.60	.35
8104245	0	.15	-1.31	12.19	20	2	12.11	.29
8 99543	4	.35	-1.10	11.43	20	2	11.44	.30
8104528	5	.59	-.47	7.11	17	2	7.11	.37
8105544	5	.71	-.07	5.17	17	3	5.16	.44
8 99334	9	.19	-1.73	9.55	16	2	9.55	.32
8105344	4	.32	-1.40	11.89	19	2	11.89	.29
8105150	9	.47	-.69	9.02	19	3	9.00	.33
8105294	5	.95	3.12	.59	19	3	.59	1.30
8103849	2	.16	-1.90	10.86	19	3	10.85	.30
8105404	4	.45	-.72	9.60	20	3	9.59	.32
8 95245	5	.56	-.51	7.98	18	3	7.97	.35
8105983	5	.40	-1.05	11.34	20	2	11.24	.30
8105129	5	.70	.00	6.17	20	3	6.14	.40
8105952	8	.47	-.85	10.09	19	3	10.09	.31
8105248	4	.55	-.49	8.39	20	3	8.39	.35
8105282	5	.40	-1.03	11.18	20	3	11.18	.30
5104114	3	.30	-.98	10.94	20	3	10.95	.30

```
      PROGRAM LINDSCO (INPUT,OUTPUT,DATA,IPOOL,TAPE1=DATA,TAPE2=IPOOL,TA
     1PE3,PUNCH)
      DIMENSION ITEM(600), A(600), B(600), C(600), INAD(300), KEY(300),
     1IREJ(300), IFORM(8), IRAW(300), INADS(300), IRESP(300), ISAD(60,25
     2), ADM(300), BDM(300), CDM(300), ISADS(300), DESC(24), NAME(2), IF
     3ORM1(8), NISS(25,2)
      INTEGER OPT1,OPT2,OPT3,OPT4
      REAL ITOT
      N=NC=0
*************************************************************************
*                                                                       *
*READ OPTIONS AND PROGRAM PARAMETER FROM INPUT FILE DATA IS ON TAPE2    *
*                                                                       *
*************************************************************************
      IPOOL=2
      READ 50, INUP,M,OPT1,OPT2,OPT3,OPT4,TS,TSS,AMAX,BMIN,BMAX,CMAX,IOM
     1IT
      READ 53, (IFORM1(I),I=1,8)
      READ (IPOOL,IFORM1) (ITEM(I),A(I),B(I),C(I),I=1,INUP)
*************************************************************************
*                                                                       *
*                                                                       *
*START READING THE SPECIFIC DATA (SPECIFIC FOR THE RUN) FROM THE INPUT  *
*INAD IS THE ITEM ID#S ADMINISTERED                                     *
*IREJ IS THE REJECTED ITEM ID S                                         *
*KEY  IS THE KEY FOR THE ITEMS IN INAD                                  *
*NSSC IS THE NUMBER OF SUBSCALES THAT WILL BE GIVEN TO THE PROGRAM      *
*NISS WILL HAVE THE #UMBER OF ITEMS IN EACH SUBSCALE TOGETHER WITH THE  *
*NAME OF THE SUBSCALES                                                  *
*ISAD IS THE ITEM ID S IN EACH SUBSCALE                                 *
*                                                                       *
*                                                                       *
*************************************************************************
      READ 51, (INAD(I),I=1,M)
      READ 52, (IREJ(I),I=1,M)
      READ 52, (KEY(I),I=1,M)
      READ 53, (IFORM(I),I=1,8)
      READ 55, DESC
      PRINT 49
      PRINT 54, INUP,M,IOMIT,OPT1,OPT2,OPT3,OPT4,TS,TSS,AMAX,BMAX,BMIN,C
     1MAX,IFORM1,IFORM,DESC
C        THOSE ITEM ID S THAT ARE IN IREJ ARE SET TO ZERO ZERO ITEM ID S
C        WILL BE SKIPPED DURING THE COMPUTATIONS
      DO 1 I=1,M
1     IF (IREJ(I).EQ.1) INAD(I)=0
C     READ SUBSCALES
      READ 48, NSSC
      IF (NSSC.EQ.0) GO TO 5
      READ 56, (NISS(I,1),NISS(I,2),I=1,NSSC)
C     READ SSC INDEX
      DO 2 JJ=1,60
      DO 2 II=1,25
2     ISAD(JJ,II)=0
      DO 3 I=1,NSSC
      NI=NISS(I,1)
      READ 51, (ISAD(J,I),J=1,NI)
3     CONTINUE
      DO 4 II=1,NSSC
      NI=NISS(II,1)
      PRINT 58, II,(ISAD(JJ,II),JJ=1,NI)
4     CONTINUE
      PRINT 49
5     CONTINUE
      IF (OPT2.EQ.1) GO TO 9
*************************************************************************
*                                                                       *
```

```
*                                                                     *      67
*IN THE NEXT DO LOOP THE ITEM PARAMETERS CORRESPONDING TO THE ITEMS    *      68
*IN INAD ARE RETRIVED FROM A,B,C, AND LOADED INTO ADM,BDM,CDM RESP.    *      69
*THE ENTRIES IN THE ADM,BDM,CDM ARE ZEROED FOR THE CASE OF ZERO ITEM   *      70
*ID IN THE INAD                                                       *      71
*                                                                     *      72
*                                                                     *      73
***********************************************************************      74
      DO 8 J=1,M                                                              75
      IF (INAD(J).NE.0) GO TO 6                                               76
      ADM(J)=BDM(J)=CDM(J)=0                                                  77
      GO TO 8                                                                 78
6     CONTINUE                                                                79
      IFOUND=0                                                                80
      DO 7 I=1,INUP                                                           81
      IF (INAD(J).NE.ITEM(I)) GO TO 7                                         82
      IFOUND=1                                                                83
      ADM(J)=A(I)                                                             84
      BDM(J)=B(I)                                                             85
      CDM(J)=C(I)                                                             86
      GO TO 8                                                                 87
7     CONTINUE                                                                88
      IF (IFOUND.EQ.0) INAD(J)=0                                              89
8     CONTINUE                                                                90
      GO TO 11                                                                91
C        THE NEXT SECTION IS USED IF OPTION 2 IS ON, IT WILL TAKE THE FI      92
C        PARAMETERS FROM THE POOL WITHOUT MAKING USE OF INAD                  93
9     DO 10 I=1,M                                                             94
      ADM(I)=A(I)                                                             95
      BDM(I)=B(I)                                                             96
      CDM(I)=C(I)                                                             97
10    CONTINUE                                                                98
C        IF OPTION 3 IS ON THE PARAMETERS A,B,C ARE CONSTRAINED WITHIN B      99
C        OF AMAX,AMIN,BMAX,BMIN,CMAX                                         100
11    IF (OPT3.EQ.0) GO TO 13                                                 101
      DO 12 I=1,M                                                             102
      IF (INAD(I).EQ.0) GO TO 12                                              103
      IF ((ADM(I).GT.AMAX).OR.(OPT3.EQ.3)) ADM(I)=AMAX                        104
      IF (BDM(I).LT.BMIN) BDM(I)=BMIN                                         105
      IF (BDM(I).GT.BMAX) BDM(I)=BMAX                                         106
      IF ((CDM(I).GT.CMAX).OR.(OPT3.EQ.2)) CDM(I)=CMAX                        107
12    CONTINUE                                                                108
13    IF (OPT1.NE.1) GO TO 15                                                 109
      DO 14 I=1,M                                                             110
      PUNCH 57, INAD(I),KEY(I),IREJ(I),ADM(I),BDM(I),CDM(I)                   111
14    CONTINUE                                                                112
15    PRINT 57, (INAD(I),KEY(I),IREJ(I),ADM(I),BDM(I),CDM(I),I=1,M)           113
      PRINT 49                                                                114
***********************************************************************      115
*                                                                     *      116
*                                                                     *      117
*READ A SUBJECT FROM TAPE1 CALCULATE THETA, LOOP BACK TO 5 ETC.       *      118
*                                                                     *      119
*                                                                     *      120
***********************************************************************      121
16    READ (1,IFORM) NAME,ID,(IRAW(I),I=1,M)                                  122
      DO 17 JJ=1,M                                                            123
17    IRESP(JJ)=0                                                             124
      IO=UNIT(1)                                                              125
C     CHECK THE  END OF FILE ON DATA FILE                                     126
      IF (IO.LE.0) GO TO 18                                                   127
      PRINT 59, IO                                                            128
18    IF (IO.EQ.0) GO TO 47                                                   129
      N=N+1                                                                   130
      ITOT=0.0                                                                131
C        SET THE ITEM ID TO ZERO FOR THE OMMISSIONS (THAT IS READ IN FRO      132
```

```
C                                                                       133
C        AND SET THE RESPONSE VECTOR TO 1 IF THE ANSWER IS CORRECT      134
C                                                                       135
      DO 21 I=1,M                                                       136
      INADS(I)=INAD(I)                                                  137
      IF (IRAW(I).EQ.IOMIT) 19,20                                       138
19    INADS(I)=0                                                        139
      GO TO 21                                                          140
20    IF ((IRAW(I).EQ.KEY(I)).AND.(IREJ(I).EQ.0)) IRESP(I)=1            141
      ITOT=ITOT+FLOAT(IRESP(I))                                         142
21    CONTINUE                                                          143
      ITINADS=0                                                         144
      DO 22 KK=1,M                                                      145
22    IF (INADS(KK).EQ.0) ITINADS=ITINADS+1                             146
      KL=M-ITINADS                                                      147
      ITOT=ITOT/FLOAT((KL))                                             148
      IF (ITOT.EQ.0.) 23,24                                             149
23    PRINT 60, NAME,ID                                                 150
      IF (OPT4.EQ.3) GO TO 26                                           151
      T=-10.0                                                           152
      SFORM=0.0                                                         153
      TINFO=0.0                                                         154
      IKL=KL                                                            155
      ITER=0                                                            156
      EXPTOT=0.0                                                        157
      WRITE (3,63) NAME,ID,ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER           158
      GO TO 32                                                          159
24    IF (ITOT.EQ.1.) 25,26                                             160
25    PRINT 61, NAME,ID                                                 161
      IF (OPT4.EQ.3) GO TO 26                                           162
      T=10.0                                                            163
      SFORM=0.0                                                         164
      TINFO=0.0                                                         165
      IKL=KL                                                            166
      ITER=0                                                            167
      EXPTOT=0.0                                                        168
      WRITE (3,63) NAME,ID,ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER           169
      GO TO 32                                                          170
*************************************************************************171
*                                                                       *172
*                                                                       *173
*     NOW THE DATA IS READY TO MAKE THE CALLS TO THE APPROPRIATE ROUTINE174
*TO ESTIMATE THE THETA.OPTION 4 WILL DETERMINE THE METHOD BY WHICH     *175
*THE THETA ESTIMATE WILL BE FOUND                                      *176
*                                                                       *177
*                                                                       *178
*************************************************************************179
26    IKL=KL                                                            180
      IF (OPT4-2) 27,28,29                                              181
27    CALL MAXLNO (IRESP,INADS,M,M,ADM,BDM,CDM,50,.005,T,SFORM,IFAIL,TIN182
     1FO,EXPTOT,ITER,SEM)                                               183
      GO TO 30                                                          184
28    CALL MAXLK (M,INADS,IRESP,M,ADM,BDM,CDM,50,.005,IFAIL,SFORM,T,TINF185
     1O,EXPTOT,ITER,SEM)                                                186
      GO TO 30                                                          187
29    T=TS                                                              188
      SFORM=TSS                                                         189
      SEM=0.0                                                           190
      CALL BAYES (M,INADS,IRESP,M,ADM,BDM,CDM,T,SFORM,TINFO,EXPTOT,ITER)191
      GO TO 31                                                          192
30    CONTINUE                                                          193
      IF (IFAIL.EQ.0) GO TO 31                                          194
      PRINT 62, NAME,ID                                                 195
C      SFORM AND T ARE SET TO -99.99 IN MAXLK IF NOT CONV               196
      NC=NC+1                                                           197
*************************************************************************198
```

```
*                                                                       *     199
*WRITE THE SUCCESFUL RESULTS TOTHE FILE TAPE3                           *     200
*                                                                       *     201
*************************************************************************     202
31    WRITE (3,63) NAME,ID,ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER,SEM             203
C        IF THERE IS NO SUBSCALE CALCULATIONS LOOP BACK TO READ A SUBJEC      204
C        OTHERWISE CONTINUE WITH THE COMPUTATIONS FOR SUBSCALES               205
32    IF (NSSC.EQ.0) GO TO 16                                                 206
      DO 46 I=1,NSSC                                                          207
      NI=NISS(I,1)                                                            208
      ITOT=0.0                                                                209
      DO 35 J=1,M                                                             210
      ISADS(J)=INADS(J)                                                       211
      DO 34 K=1,NI                                                            212
      IF (ISAD(K,I)-INADS(J)) 34,33,34                                        213
33    ITOT=ITOT+FLOAT(IRESP(J))                                               214
      GO TO 35                                                                215
34    CONTINUE                                                                216
      ISADS(J)=0                                                              217
35    CONTINUE                                                                218
      ITISADS=0                                                               219
      DO 36 KK=1,M                                                            220
36    IF (ISADS(KK).NE.0) ITISADS=ITISADS+1                                   221
      KL=ITISADS                                                              222
      ITOT=ITOT/FLOAT(KL)                                                     223
      IF (ITOT.EQ.0.0) 37,38                                                  224
37    PRINT 66, NAME,ID,NISS(I,2)                                             225
      IF (OPT4.EQ.3) GO TO 40                                                 226
      T=-10.00                                                                227
      TINFO=EXPTOT=0.0                                                        228
      ITER=0                                                                  229
      SFORM=0.0                                                               230
      WRITE (3,64) NAME,ID,NISS(I,2),ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER       231
      GO TO 46                                                                232
38    IF (ITOT.EQ.1.0) 39,40                                                  233
39    PRINT 67, NAME,ID,NISS(I,2)                                             234
      IF (OPT4.EQ.3) GO TO 40                                                 235
      T=10.0                                                                  236
      TINFO=EXPTOT=0.0                                                        237
      ITER=0                                                                  238
      SFORM=0.0                                                               239
      WRITE (3,64) NAME,ID,NISS(I,2),ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER       240
      GO TO 46                                                                241
40    IKL=KL                                                                  242
      IF (OPT4-2) 41,42,43                                                    243
41    CALL MAXLNO (IRESP,ISADS,M,M,ADM,BDM,CDM,50,.01,T,SFORM,IFAIL,TINF      244
     1O,EXPTOT,ITER,SEM)                                                      245
      GO TO 44                                                                246
42    CALL MAXLK (M,ISADS,IRESP,M,ADM,BDM,CDM,50,.01,IFAIL,SFORM,T,TINFO      247
     1,EXPTOT,ITER,SEM)                                                       248
      GO TO 44                                                                249
43    T=TS                                                                    250
      SFORM=TSS                                                               251
      SEM=0.0                                                                 252
      CALL BAYES (M,ISADS,IRESP,M,ADM,BDM,CDM,T,SFORM,TINFO,EXPTOT,ITER)      253
      GO TO 45                                                                254
44    IF (IFAIL.EQ.0) GO TO 45                                                255
      PRINT 68, NAME,ID,NISS(I,2)                                             256
45    WRITE (3,64) NAME,ID,NISS(I,2),ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER,      257
     1SEM                                                                     258
46    CONTINUE                                                                259
      GO TO 16                                                                260
47    PRINT 65, N,NC                                                          261
      STOP                                                                    262
C                                                                             263
C                                                                             264
```

```
C                                                                       265
48     FORMAT (I5)                                                      266
49     FORMAT (1H1)                                                     267
50     FORMAT (2I4,X,4I1,6F5.2,2X,I2)                                   268
51     FORMAT (16I5)                                                    269
52     FORMAT (80I1)                                                    270
53     FORMAT (8A10)                                                    271
54     FORMAT (T50,*LINDSCO*,/,T50*=======*,//////,T20,*LINEAR DICHOTOMUS  272
      1 SCORING WITH THREE PARAMETER MODELS*,////,T40*PSYCHOMETRIC METHOD  273
      2S PROGRAM*,/,T40*DEPARTMENT OF PSYCHOLOGY*,/,T40*UNIVERSITY OF MIN  274
      3NESOTA*,/,T40*MPLS. MINN. 55455*,/,///,T20*INUP*,T27*=*I5,/,T20,*M  275
      4MAX*,T27*=*I5,/,T20,*IOMIT*,T27*=*,I5,/,T20*OPT1*,T27*=*I5,/,T20*O  276
      5PT2*,T27*=*I5,/,T20*OPT3*,T27*=*I5,/,T20*OPT4*,T27*=*,I5,/,T20*TS*  277
      6,T27*=*F5.2,/,T20*TSS*,T27*=*F5.2,/,T20*AMAX*,T27*=*F5.2,/,T20,*BM  278
      7AX*,T27,*=*F5.2,/,T20*BMIN*,T27,*=*F5.2,/,T20*CMAX*,T27,*=*F5.2,/,  279
      8T20,*VARIABLE FORMAT FOR POOL=*8A10,/,T20,*VARIABLE FORMAT FOR DAT  280
      9A=*8A10,/,T20,8A10,/,T20,8A10,/,T20,8A10,/)                         281
55     FORMAT (8A10)                                                    282
56     FORMAT (2I5)                                                     283
57     FORMAT (*1*,29X,*ITEM ID S*2X,*KEYS*2X,*REJECTIONS*,4X,*A*,9X,*B*,  284
      19X,*C*,/,60(/30X,I5,5X,I3,5X,I3,1X,F10.2,1X,F10.2,1X,F10.2))        285
58     FORMAT (///,40X,*ITEMS IN SUBSCALE NO=*I3,/,10(20I6,/))          286
59     FORMAT (10X,*PARITY ERROR ON TAPE*,90X,I2)                       287
60     FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS  NO ANSWERS *,*CORRE  288
      1CT IN TOTAL SCALE*)                                                 289
61     FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS ALL ANSWERS *,*CORRE  290
      1CT IN TOTAL SCALE*)                                                 291
62     FORMAT (10X,*COMPUTATIONAL PROBLEMS WITH SUBJECT =*,2A10,* ID =*,A  292
      19,*  IN TOTAL TE*)                                                 293
63     FORMAT (X,2A10,A9,* T*,F5.2,2F7.2,I4,2F7.2,I4,F7.2)              294
64     FORMAT (X,2A10,A9,I2,F5.2,2F7.2,I4,2F7.2,I4,F7.2)                295
65     FORMAT (/10X,*CASES  READ=*,I5,* CASES NOT  CONVERGED=*,I5)      296
66     FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS NO  ANSWERS *,*CORRE  297
      1CT IN SUBSCALE *,I5)                                                298
67     FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS ALL ANSWERS *,*CORRE  299
      1CT IN SUBSCALE *,I5)                                                300
68     FORMAT (10X,*MAXIMUM LIKELIHOOD ESTIMATION DOES NOT CONVERGE*,*FOR  301
      1 THE SUBJECT = *,2A10,* ID = *,A9,*  ON SUBSCALE *,I5)              302
      END                                                              303
      SUBROUTINE BAYES (M,ITM,RESP,N,A,B,C,BTHET,BVAR,TINFO,EXPTOT,ITER)    1
      INTEGER RESP(M),ITM(M)                                           2
      REAL A(N),B(N),C(N)                                              3
      DO 1 I=1,M                                                       4
      IF (ITM(I).EQ.0) GO TO 1                                         5
      CALL BSCOR (BTHET,BVAR,B(I),A(I),C(I),RESP(I))                   6
1     CONTINUE                                                         7
      CALL NOSTAT (M,ITM,A,B,C,BTHET,TINFO,EXPTOT)                     8
      ITER=0                                                           9
      RETURN                                                          10
      END                                                             11
      SUBROUTINE BSCOR (BTHET,BVAR,DIF,DIS,GUESP,IRESP)                1
      D=(DIF-BTHET)/SQRT(2.0*(1.0/DIS**2+BVAR))                        2
      ERFD=ERFNP(D)                                                    3
      EDSQ=EXP(D**2)                                                   4
      IF (EDSQ.EQ.0.0) RETURN                                          5
      EDSQI=1.0/EDSQ                                                   6
      XKINV=0.5*(1.0-ERFD)                                             7
      XLINV=GUESP+(1.0-GUESP)*XKINV                                    8
      IF ((XLINV.EQ.0.0).OR.(XKINV.EQ.0.0)) RETURN                     9
      XL=1.0/XLINV                                                    10
      IF (IRESP.NE.1) GO TO 1                                         11
      S=0.398942*(SQRT(BVAR)/SQRT(1.0+(1.0/DIS**2)/BVAR))*(1.0/XKINV)*ED  12
      1SQI                                                            13
      T=1.0-1.772454*D*EDSQ*(1.0-ERFD)                                14
      BTHET=BTHET+(1.0-GUESP)*XKINV*XL*S                              15
      BVAR=BVAR-(1.0-GUESP)*XKINV*XL*S**2*(T-GUESP*XL)                16
```

```
      RETURN                                                            17
1     BTHET=BTHET-0.797885*(BVAR/SQRT(1.0/DIS**2+BVAR))*EDSQI*(1.0/(1.0+ 18
     1ERFD))                                                            19
      PART1=1.128379/(1.0+(1.0/DIS**2)*(1.0/BVAR))                      20
      PART2=1.0/(EDSQ*(1.0+ERFD))**2                                    21
      PART3=0.564190+0*EDSQ*(1.0+ERFD)                                  22
      BVAR=BVAR*(1.0-PART1*PART2*PART3)                                 23
      RETURN                                                            24
      END                                                               25
      REAL FUNCTION ERFNP (X)                                            1
      DATA A1/0.254830/                                                  2
      DATA A2/-0.284497/                                                 3
      DATA A3/1.421414/                                                  4
      DATA A4/-1.453152/                                                 5
      DATA A5/1.061405/                                                  6
      DATA P/0.327591/                                                   7
      ERFNP=0.0                                                          8
      IF (X.EQ.0.0) RETURN                                               9
      ES=SIGN(1.0,X)                                                    10
      Y=ABS(X)                                                          11
      IF (Y.LT.6.0) GO TO 1                                             12
      ERFNP=ES                                                          13
      RETURN                                                            14
1     Y2=Y*Y                                                            15
      T=1.0/(1.0+P*Y)                                                   16
      AT=((A1+(A2+(A3+(A4+A5*T)*T)*T)*T)*T)                             17
      EAT=AT/EXP(Y2)                                                    18
      ERFNP=(1.0-EAT)*ES                                                19
      RETURN                                                            20
      END                                                               21
      SUBROUTINE MAXLK (M,ITM,RESP,N,A,B,C,MAX,EPS,IFAIL,SDRV,THETA,TINF  1
     1O,EXPTOT,NUMITS,SEM)                                              2
      EXTERNAL FDDLOG,SDDLOG                                             3
      INTEGER RESP(M)                                                    4
      DIMENSION A(N), B(N), C(N), ITM(M)                                 5
C***  USES MAXIMUM LIKELIHOOD LOGISTIC SCORING ALGORITHM AND RESPONSE    6
C***  MODEL                                                              7
C***  BISECTION IS USED TO PROVIDE THE INITIAL GUESS FOR THE             8
C***  NEWTON-RAPHSON METHOD                                              9
      CALL BISECT (FDDLOG,RESP,A,B,C,M,ITM,5,GUESS)                     10
C                                                                       11
      CALL NEWTRAP (FDDLOG,SDDLOG,RESP,A,B,C,M,ITM,MAX,EPS,NUMITS,GUESS, 12
     1THETA,SDRV,IFAIL)                                                 13
C                                                                       14
      IF (IFAIL.EQ.1) 1,2                                               15
C***  NEWTON RAPHSON DID NOT CONVERGE                                   16
1     CALL NWTERR (THETA,SDRV,SEM,TINFO,EXPTOT)                         17
      RETURN                                                            18
C                                                                       19
2     CALL LGSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                      20
      SEM=1.0/SQRT(ABS(SDRV))                                           21
      RETURN                                                            22
      END                                                               23
      FUNCTION FDDLOG (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA XMAX,XMIN/200.0,-200.0/                                       4
      DATA D,RIGHT/1.7,1/                                                5
C***  CALCULATES FIRST DERIVATIVE OF LOG-LIKELIHOOD FUNCTION OF A        6
C***  RESPONSE VECTOR FOR THE LOGISTIC MODEL                             7
      SUM=0.0                                                            8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0) GO TO 1                                          10
      X=D*A(I)*(THETA-B(I))                                             11
      IF (X.LT.XMIN) X=XMIN                                             12
      IF (X.GT.XMAX) X=XMAX                                             13
```

```
      EXF=EXP(X)                                                        14
      AE=A(I)*EXF                                                       15
      SUM=SUM-AE/(EXF+1.0)                                              16
      IF (RESP(I).NE.RIGHT) GO TO 1                                     17
      CE=C(I)+EXF                                                       18
      SUM=SUM+AE/CE                                                     19
1     CONTINUE                                                          20
      FDDLOG=-1.7*SUM                                                   21
      RETURN                                                            22
      END                                                               23
      FUNCTION SDDLOG (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION ITM(M), A(M), B(M), C(M)                                 3
      DATA XMAX,XMIN/200.0,-200.0/                                       4
      DATA D,RIGHT/1.7,1/                                                5
C***  CALCULATES SECOND DERIVATIVE OF LOG-LIKELIHOOD FUNCTION            6
C***  OF A RESPONSE VECTOR FOR THE LOGISTIC MODEL                        7
      SUM=0.0                                                            8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0) GO TO 1                                          10
      X=D*A(I)*(THETA-B(I))                                             11
      IF (X.LT.XMIN) X=XMIN                                             12
      IF (X.GT.XMAX) X=XMAX                                             13
      EXF=EXP(X)                                                        14
      AE=A(I)*EXF                                                       15
      SUM=SUM-A(I)*AE/((1.0+EXF)*(1.0+EXF))                             16
      IF (RESP(I).NE.RIGHT) GO TO 1                                     17
      CE=C(I)+EXF                                                       18
      SUM=SUM+A(I)*C(I)*AE/(CE*CE)                                      19
1     CONTINUE                                                          20
      SDDLOG=-2.89*SUM                                                  21
      RETURN                                                            22
      END                                                               23
      SUBROUTINE BISECT (F1,RESP,A,B,C,M,ITM,NITER,BMID)                 1
      INTEGER RESP(M)                                                    2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
C***  CALCULATES APPROXIMATE ROOT OF F1 BY BISECTION;                    4
C***  BISECTING NITER (NUMBER OF ITERATIONS) TIMES.                      5
C***  BMID IS BEST CURRENT GUESS AT ROOT THETA                           6
C                                                                        7
C***  INITIALIZE LEFT BOUND AND F1(BOUND) AND RIGHT BOUND F1(BOUND)      8
      BL=-5.0                                                            9
      BR=5.0                                                            10
      BMID=0.0                                                          11
      TL=F1(RESP,ITM,A,B,C,M,BL)                                        12
      TR=F1(RESP,ITM,A,B,C,M,BR)                                        13
C***  TEST FOR NO ROOT IN INTERVAL--RETURN IF NO  SOLUTION              14
      IF ((TL*TR).GT.0.0) RETURN                                        15
C                                                                       16
C***  NOW CALCULATE BISECTIONS NITER TIMES                              17
      DO 3 I=1,NITER                                                    18
      TMID=F1(RESP,ITM,A,B,C,M,BMID)                                    19
      IF ((TMID*TL).GT.0.0) GO TO 1                                     20
C***  REPLACE RIGHT BOUND WITH BMID                                     21
      BR=BMID                                                           22
      GO TO 2                                                           23
C***  REPLACE LEFT BOUND WITH BMID                                      24
1     TL=TMID                                                           25
      BL=BMID                                                           26
C***  FIND NEW MIDPOINT BMID                                            27
2     BMID=(BL+BR)/2.0                                                  28
3     CONTINUE                                                          29
      RETURN                                                            30
      END                                                               31
      SUBROUTINE NEWTRAP (F1,F2,RESP,A,B,C,M,ITM,NITER,EPS,NUMITS,GUESS,  1
     1THETA,SDRV,IFAIL)                                                   2
```

```
      INTEGER RESP(M)                                                    3
      DIMENSION A(M), B(M), C(M), ITM(M)                                 4
C***  CALCULATES THE ROOT OF F1 GIVEN ITS FIRST DERIVATIVE F2            5
C***  AND AN INITIAL GUESS USING NEWTON-RAPHSON METHOD                   6
C***  THETA IS APPR. TO THE ROOT; SDRV IS F2(THETA)                      7
      NUMITS=0                                                           8
      THETA=GUESS                                                        9
C***  LOOP UNTIL ERR<EPS OR NUMBER OF ITERATIONS BECOMES TOO LARGE      10
1     FDRV=F1(RESP,ITM,A,B,C,M,THETA)                                   11
      SDRV=F2(RESP,ITM,A,B,C,M,THETA)                                   12
      ERR=FDRV/SDRV                                                     13
      THETA=THETA-ERR                                                   14
      NUMITS=NUMITS+1                                                   15
C***  EXIT LOOP CRITERION                                               16
      IF ((NUMITS.LT.NITER).AND.(ABS(ERR).GT.EPS)) GO TO 1              17
C***  END LOOP.  TEST FOR CONVERGENCE AND SET IFAIL                     18
      IFAIL=0                                                           19
      IF (ABS(ERR).LT.EPS) RETURN                                       20
C                                                                       21
C***  NEWTON RAPHSON METHOD DOES NOT CONVERGE                           22
      IFAIL=1                                                           23
      RETURN                                                            24
      END                                                               25
      SUBROUTINE NWTERR (THETA,SFORM,SEM,TINFO,EXPTOT)                   1
C***  SETS ERROR VALUES FOR THE CASE IN WHICH NEWTON RAPHSON FAILS       2
C***  TO CONVERGE                                                        3
      THETA=-99.99                                                       4
      SFORM=-99.99                                                       5
      SEM=-99.99                                                         6
      TINFO=-99.99                                                       7
      EXPTOT=-99.99                                                      8
      RETURN                                                             9
      END                                                               10
      SUBROUTINE LGSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                 1
      DIMENSION A(M), B(M), C(M), ITM(M)                                 2
      DATA XMAX,XMIN/12.0,-12.0/                                         3
      TINFO=0.0                                                          4
      EXPTOT=0.0                                                         5
      KOUNT=0                                                            6
C                                                                        7
      DO 1 I=1,M                                                         8
      IF (ITM(I).EQ.0) GO TO 1                                           9
      KOUNT=KOUNT+1                                                     10
      ARGU=-1.7*A(I)*(THETA-B(I))                                       11
      IF (ARGU.GT.XMAX) ARGU=XMAX                                       12
      IF (ARGU.LT.XMIN) ARGU=XMIN                                       13
      P=C(I)+(1.0-C(I))*(1.0/(1.0+EXP(ARGU)))                           14
      Q=1.0-P                                                           15
      EARG=EXP(-ARGU)                                                   16
      PPRIME=EARG/((1.0+EARG)*(1.0+EARG))                               17
      PPRIME=PPRIME*(1.0-C(I))*A(I)*1.7                                 18
      TINFO=TINFO+(PPRIME*PPRIME)/(P*Q)                                 19
      EXPTOT=EXPTOT+P                                                   20
1     CONTINUE                                                          21
      EXPTOT=EXPTOT/FLOAT(KOUNT)                                        22
      RETURN                                                            23
      END                                                               24
      SUBROUTINE MAXLNO (RESP,ITM,M,N,A,B,C,MAX,EPS,THETA,SDRV,IFAIL,TIN  1
     1FO,EXPTOT,NUMITS,SEM)                                             2
      EXTERNAL FDNOGV,SDNOGV                                             3
      INTEGER RESP(M)                                                    4
      DIMENSION ITM(N), A(N), B(N), C(N)                                 5
C***  USES MAXIMUM LIKELIHOOD NORMAL OGIVE SCORING ALGORITHM AND         6
C***  RESPONSE VECTOR                                                    7
C***  BISECTION IS USED TO PROVIDE THE INITIAL GUESS FOR THE             8
C***  NEWTON RAPHSON METHOD                                              9
```

```
C                                                                       10
      CALL BISECT (FDNOGV,RESP,A,B,C,M,ITM,5,GUESS)                     11
C                                                                       12
      CALL NEWTRAP (FDNOGV,SDNOGV,RESP,A,B,C,M,ITM,MAX,EPS,NUMITS,GUESS,13
     1THETA,SDRV,IFAIL)                                                 14
      IF (IFAIL.EQ.1) 1,2                                               15
C***  NEWTON RAPHSON DID NOT CONVERGE                                   16
1     CALL NWTERR (THETA,SDRV,SEM,TINFO,EXPTOT)                         17
      RETURN                                                            18
C                                                                       19
2     CALL NOSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                      20
      SDRV=ABS(SDRV)                                                    21
      SEM=1.0/SQRT(SDRV)                                                22
      RETURN                                                            23
      END                                                               24
      FUNCTION FDNOGV (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA PI,RIGHT/3.141592,1/                                          4
      DATA XMAX,XMIN/7.0,-7.0/                                           5
C***  CALCULATES FIRST DERIVATIVE OF LOG-LIKELIHOOD FUNCTION OF          6
C***  A RESPONSE VECTOR FOR THE NORMAL OGIVE MODEL                       7
C                                                                        8
      SUM=0.0                                                            9
      ROOTPI=1.0/SQRT(2.0*PI)                                           10
      DO 2 I=1,M                                                        11
      IF (ITM(I).EQ.0) GO TO 2                                          12
      TEMP=A(I)*(THETA-B(I))                                            13
      IF (TEMP.GT.XMAX) TEMP=XMAX                                       14
      IF (TEMP.LT.XMIN) TEMP=XMIN                                       15
      X=-(TEMP*TEMP)/2.0                                                16
      DNMRAT=ROOTPI*A(I)*(1.0-C(I))*EXP(X)                              17
      DENOM=C(I)+(1.0-C(I))*CDFN(TEMP)                                  18
      IF (RESP(I).EQ.RIGHT) GO TO 1                                     19
      DENOM=-(1.0-DENOM)                                                20
1     SUM=SUM+(DNMRAT/DENOM)                                            21
2     CONTINUE                                                          22
      FDNOGV=SUM                                                        23
      RETURN                                                            24
      END                                                               25
      FUNCTION SDNOGV (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA PI,RIGHT/3.141592,1/                                          4
      DATA XMAX,XMIN/7.0,-7.0/                                           5
C***  CALCULATES SECOND DERIVATIVE OF LOG-LIKELIHOOD FUNCTION            6
C***  OF A RESPONSE VECTOR FOR THE NORMAL OGIVE MODEL                    7
C                                                                        8
      SUM=0.0                                                            9
      ROOTPI=1.0/SQRT(2.0*PI)                                           10
      DO 2 I=1,M                                                        11
      IF (ITM(I).EQ.0) GO TO 2                                          12
      TEMP1=A(I)*(THETA-B(I))                                           13
      IF (TEMP1.GT.XMAX) TEMP1=XMAX                                     14
      IF (TEMP1.LT.XMIN) TEMP1=XMIN                                     15
      X=-TEMP1*TEMP1/2.0                                                16
      TEMP2=ROOTPI*(1.0-C(I))*A(I)*EXP(X)                               17
      FIRNUM=TEMP2*TEMP2                                                18
      SECNUM=TEMP2*A(I)*TEMP1                                           19
      SDENOM=C(I)+(1.0-C(I))*CDFN(TEMP1)                                20
      FDENOM=SDENOM*SDENOM                                              21
      IF (RESP(I).EQ.RIGHT) GO TO 1                                     22
      FDENOM=(1.0-SDENOM)*(1.0-SDENOM)                                  23
      SDENOM=-(1.0-SDENOM)                                              24
1     SUM=SUM-(FIRNUM/FDENOM)-(SECNUM/SDENOM)                           25
2     CONTINUE                                                          26
```

```
      SDNOGV=SUM                                                        27
      RETURN                                                            28
      END                                                               29
      SUBROUTINE NOSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                 1
      DIMENSION A(M), B(M), C(M), ITM(M)                                 2
      DATA PI/3.141592/                                                  3
      DATA XMAX,XMIN/7.0,-7.0/                                           4
      TINFO=0.0                                                          5
      EXPTOT=0.0                                                         6
      KOUNT=0                                                            7
C                                                                        8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0) GO TO 1                                          10
      KOUNT=KOUNT+1                                                     11
      TEMP=A(I)*(THETA-B(I))                                            12
      IF (TEMP.GT.XMAX) TEMP=XMAX                                       13
      IF (TEMP.LT.XMIN) TEMP=XMIN                                       14
      P=C(I)+(1.0-C(I))*CDFN(TEMP)                                      15
      Q=1.0-P                                                           16
      TEMP=-TEMP*TEMP/2.0                                               17
      PPRIME=(1.0/SQRT(2.0*PI))*(1.0-C(I))*A(I)*EXP(TEMP)               18
      TINFO=TINFO+(PPRIME*PPRIME)/(P*Q)                                 19
      EXPTOT=EXPTOT+P                                                   20
1     CONTINUE                                                          21
      EXPTOT=EXPTOT/KOUNT                                               22
      RETURN                                                            23
      END                                                               24
```

```
      PROGRAM ADADSCO (INPUT,OUTPUT,DATA,IPOOL,TAPE1=DATA,TAPE2=IPOOL,TA
     1PE3)
      DIMENSION ITEM(600), A(600), B(600), C(600), KEY(600), IREJ(80), I
     1FORM2(8), IRAW(80), INADS(80), IRESP(80), ADM(80), BDM(80), CDM(80
     2), DESC(24), NAME(2), IFORM1(8)
      INTEGER OPT1,OPT2,OPT3,OPT4
      REAL ITOT
      N=NC=0
C*************************************************************************
C                                                                        *
CREAD OPTIONS AND PROGRAM PARAMETER FROM INPUT FILE, DATA IS ON TAPE2    *
C                                                                        *
C*************************************************************************
      IPOOL=2
      READ 20, INUP,MMAX,OPT1,OPT2,OPT3,OPT4,TS,TSS,AMAX,BMIN,BMAX,CMAX,
     1IFLAG,IOMIT
      READ 22, (IFORM1(I),I=1,8)
C        IFORM1  IS THE VARIABLE FORMAT FOR THE ITEM POOL
      READ (IPOOL,IFORM1) (ITEM(I),A(I),B(I),C(I),KEY(I),I=1,INUP)
C*************************************************************************
C                                                                       *
C                                                                       *
CSTART READING THE SPECIFIC DATA (SPECIFIC FOR THE RUN) FROM THE INPUT  *
CINAD IS THE ITEM ID#S ADMINISTERED                                     *
CIREJ IS THE REJECTED ITEM ID S                                         *
CIRESP IS THE RESPONSE VECTOR                                           *
C                                                                       *
C                                                                       *
C*************************************************************************
      READ 21, MNUM,(IREJ(I),I=1,MNUM)
      READ 22, (IFORM2(I),I=1,8)
C          IFORM2 IS THE VARIABLE FORMAT FOR THE SUBJECT DATA
C
      READ 24, DESC
      PRINT 23, INUP,MMAX,IOMIT,IFLAG,OPT1,OPT2,OPT3,OPT4,TS,TSS,AMAX,BM
     1AX,BMIN,CMAX,IFORM1,IFORM2,DESC
C*************************************************************************
C                                                                       *
C                                                                       *
CREAD A SUBJECT FROM TAPE1 CALCULATE THETA, LOOP BACK TO 5 ETC.         *
C                                                                       *
C                                                                       *
C*************************************************************************
1     READ (1,IFORM2) ID,NAME,M,(IRESP(I),I=1,MMAX),(INADS(I),I=1,MMAX)
      M=MIN0(M,MMAX)
      IO=UNIT(1)
C     CHECK THE  END OF FILE ON DATA FILE
C
      IF (IO.LE.0) GO TO 2
      PRINT 25, IO
2     IF (IO.EQ.0) GO TO 19
      N=N+1
      ITOT=0.0
C        SET THE ITEM ID TO ZERO FOR THE OMMISSIONS (THAT IS READ IN FRO
C
C        AND SET THE RESPONSE VECTOR TO 1 IF THE ANSWER IS CORRECT
      DO 5 I=1,M
C        SET THE ITEM IDS IN IREEJ TO ZERO
      IF (MNUM.EQ.0) GO TO 4
      DO 3 IJ=1,MNUM
      IF (INADS(I).NE.IREJ(IJ)) GO TO 3
      INADS(I)=0
      GO TO 5
3     CONTINUE
4     CONTINUE
      IF (IRESP(I).EQ.IOMIT) INADS(I)=0
```

```
      IF (INADS(I).EQ.0) GO TO 5                                          67
      CALL SEARCH (INUP,INADS(I),ADM(I),BDM(I),CDM(I),KKE,ITEM,A,B,C,KEY  68
     1,ID)                                                                69
C     IF THE FLAG IFLAG IS NOT ZERO IT IS TAKEN TO BE THE DUMMY KEY       70
C     IF IT IS ZERO THEN KEY IS READ FROM THE POOL AN LEFT IN KKE         71
      IF (IFLAG.NE.0) KKE=IFLAG                                           72
      IRES=0                                                              73
      IF (IRESP(I).EQ.KKE) IRES=1                                         74
      IRESP(I)=IRES                                                       75
5     CONTINUE                                                            76
C*************************************************************************77
C                                                                        *  78
C                                                                        *  79
CIN THE NEXT DO LOOP THE ITEM PARAMETERS CORRESPONDING TO THE ITEMS      *  80
CIN INAD ARE RETRIVED FROM A,B,C, AND LOADED INTO ADM,BDM,CDM RESP.      *  81
CTHE ENTRIES IN THE ADM,BDM,CDM ARE ZEROED FOR THE CASE OF ZERO ITEM     *  82
CID IN THE INAD                                                          *  83
C                                                                        *  84
C                                                                        *  85
C*************************************************************************86
C        IF OPTION 3 IS ON THE PARAMETERS A,B,C ARE CONSTRAINED WITHIN B     87
C        OF AMAX,AMIN,BMAX,BMIN,CMAX                                         88
      IF (OPT3.EQ.0) GO TO 7                                              89
      DO 6 I=1,M                                                          90
      IF (INADS(I).EQ.0) GO TO 6                                          91
      IF ((ADM(I).GT.AMAX).OR.(OPT3.EQ.3)) ADM(I)=AMAX                    92
      IF (BDM(I).LT.BMIN) BDM(I)=BMIN                                     93
      IF (BDM(I).GT.BMAX) BDM(I)=BMAX                                     94
      IF ((CDM(I).GT.CMAX).OR.(OPT3.EQ.2)) CDM(I)=CMAX                    95
6     CONTINUE                                                            96
C        OPTION 1 WILL PRINT THE SPECIFIC DATA IF ITS ON                  97
C                                                                         98
7     CONTINUE                                                            99
      IF ((OPT1.EQ.0).OR.(OPT1.GT.10)) GO TO 8                           100
      OPT1=OPT1+1                                                        101
      PRINT 26, ID,NAME,(INADS(II),IRESP(II),ADM(II),BDM(II),CDM(II),II= 102
     11,M)                                                               103
8     CONTINUE                                                           104
      ITINADS=0                                                          105
      DO 9 KK=1,M                                                        106
      ITOT=ITOT+IRESP(KK)                                                107
9     IF (INADS(KK).EQ.0) ITINADS=ITINADS+1                              108
      KL=M-ITINADS                                                       109
      ITOT=ITOT/FLOAT((KL))                                              110
      IF (ITOT.EQ.0) 10,11                                               111
10    PRINT 27, NAME,ID                                                  112
      IF (OPT4.EQ.3) GO TO 13                                            113
      T=-10.0                                                            114
      SFORM=0.0                                                          115
      TINFO=0.0                                                          116
      EXPTOT=0.0                                                         117
      SEM=0.0                                                            118
      ITER=0                                                             119
      IKL=KL                                                             120
      GO TO 18                                                           121
11    IF (ITOT.EQ.1.0) 12,13                                             122
12    PRINT 31, NAME,ID                                                  123
      IF (OPT4.EQ.3) GO TO 13                                            124
      T=10.0                                                             125
      SFORM=0.0                                                          126
      TINFO=0.0                                                          127
      EXPTOT=0.0                                                         128
      SEM=0.0                                                            129
      ITER=0                                                             130
      IKL=KL                                                             131
      GO TO 18                                                           132
```

```
C*************************************************************************      133
C                                                                       *      134
C                                                                       *      135
C     NOW THE DATA IS READY TO MAKE THE CALLS TO THE APPROPRIATE ROUTINE      136
CTO ESTIMATE THE THETA. OPTIN 4 WILL DETERMINE THE METHOD BY WHICH      *      137
CTHE THETA ESTIMATE WILL BE FOUND                                       *      138
C                                                                       *      139
C                                                                       *      140
C*************************************************************************      141
13    IKL=KL                                                                   142
      IF (OPT4-2) 14,15,16                                                     143
14    CALL MAXLNO (IRESP,INADS,M,M,ADM,BDM,CDM,50,.005,T,SFORM,IFAIL,TIN       144
     1FO,EXPTOT,ITER,SEM)                                                      145
      GO TO 17                                                                 146
15    CALL MAXLK (M,INADS,IRESP,M,ADM,BDM,CDM,50,.005,IFAIL,SFORM,T,TINF       147
     10,EXPTOT,ITER,SEM)                                                       148
C                                                                              149
      GO TO 17                                                                 150
16    T=TS                                                                     151
      SFORM=TSS                                                                152
      CALL BAYES (M,INADS,IRESP,M,ADM,BDM,CDM,T,SFORM,TINFO,EXPTOT)            153
      ITER=0                                                                   154
      SEM=0.0                                                                  155
      GO TO 18                                                                 156
17    IF (IFAIL.EQ.0) GO TO 18                                                 157
      PRINT 28, NAME,ID                                                        158
C      SFORM AND T ARE SET TO -99.99 IN MAXLK IF NOT CONV                      159
      NC=NC+1                                                                  160
C*************************************************************************      161
C                                                                       *      162
CWRITE THE SUCCESFUL RESULTS TOTHE FILE TAPE3                           *      163
C                                                                       *      164
C*************************************************************************      165
18    WRITE (3,29) NAME,ID,ITOT,T,SFORM,IKL,TINFO,EXPTOT,ITER,SEM              166
      GO TO 1                                                                  167
19    PRINT 30, N,NC                                                           168
C                                                                              169
      STOP                                                                     170
C                                                                              171
C                                                                              172
20    FORMAT (2I4,X,4I1,6F5.2,I2,I2)                                           173
21    FORMAT (16I5)                                                            174
22    FORMAT (8A10)                                                            175
23    FORMAT (T50,*ADADSCO*,/,T50*=======*,//////,T20,*ADAPTIVE DICHOTOM       176
     1US SCORING WITH THREE PARAMETER MODELS*,////,T40*PSYCHOMETRIC METH       177
     2ODS PROGRAM*,/,T40*DEPARTMENT OF PSYCHOLOGY*,/,T40*UNIVERSITY OF M       178
     3INNESOTA*,/,T40*MPLS. MINN. 55455*,/,///,T20*INUP*,T27*=*I5,/,T20,       179
     4*MMAX*,T27*=*I5,/,T20,*IOMIT*,T27*=*,I5,/,T20*IFLAG*,T27*=*,I5,/,T       180
     520*OPT1*,T27*=*I5,/,T20*OPT2*,T27*=*I5,/,T20*OPT3*,T27*=*I5,/,T20*       181
     6OPT4*,T27*=*,I5,/,T20*TS*,T27*=*F5.2,/,T20*TSS*,T27*=*F5.2,/,T20*A       182
     7MAX*,T27*=*F5.2,/,T20,*BMAX*,T27,*=*F5.2,/,T20*BMIN*,T27,*=*F5.2,/       183
     8,T20*CMAX*,T27,*=*F5.2,/,T20,*VARIABLE FORMAT FOR POOL=*8A10,/,T20       184
     9,*VARIABLE FORMAT FOR DATA=*8A10,/,T20,8A10,/,T20,8A10,/,T20,8A10,       185
     8/)                                                                       186
24    FORMAT (8A10)                                                            187
25    FORMAT (10X,*PARITY ERROR ON TAPE*,90X,I2)                               188
26    FORMAT (10X,A9,2A10,/,(1X,I4,2X,I2,2X,3F10.2))                           189
27    FORMAT (10X,* SUBJECT *,2A10,*ID= *,A9,*HAS NO RIGHT ANSWERS*)           190
28    FORMAT (10X,* MAXIMUM LIKELIHOOD ESTIMATION DOES NOT CONVERGE*,* F       191
     10R THE SUBJECT = *,2A10,*  ID= *,A9)                                     192
29    FORMAT (X,2A10,A9,F5.2,2F7.2,I4,2F7.2,I4,F7.2)                           193
30    FORMAT (10X,* CASES  READ=*,I5,* CASES NOT  CONVERGED=*,I5)              194
31    FORMAT (10X,* SUBJECT *,2A10,*ID= *,A9,*HAS ALL ANSWERS RIGHT*)          195
      END                                                                      196
      SUBROUTINE SEARCH (INUP,ID,A,B,C,KEY,ITM,AP,BP,CP,KEYP,IDNUM)             1
      DIMENSION AP(INUP), BP(INUP), CP(INUP), KEYP(INUP), ITM(INUP)             2
```

```
      INTEGER FLAG                                                          3
      DO 1 I=1,INUP                                                         4
      IF (ID.NE.ITM(I)) GO TO 1                                             5
      A=AP(I)                                                               6
      B=BP(I)                                                               7
      C=CP(I)                                                               8
      KEY=KEYP(I)                                                           9
      CALL PCHECK (A,B,C,ID,IDNUM)                                         10
      RETURN                                                               11
1     CONTINUE                                                             12
      PRINT 2, ID,IDNUM                                                    13
C                                                                          14
      ID=0                                                                 15
      RETURN                                                               16
C                                                                          17
C                                                                          18
C                                                                          19
2     FORMAT (10X,* ITEM =*,I4,* IS NOT IN THE POOL FOR SUBJECT ID =*,A9   20
     1)                                                                    21
      END                                                                  22
      SUBROUTINE PCHECK (A,B,C,ID,IDNUM)                                    1
C***  CHECK WHETHER OR NOT ITEM PARAMETERS ARE VALID                        2
C***  IF NOT, ERROR MESSAGE IS PRINTED                                      3
C                                                                           4
      IF (A.LE.0.0) PRINT 1, ID,A,IDNUM                                     5
      IF ((-5.0.GT.B).OR.(B.GT.5.0)) PRINT 2, ID,B,IDNUM                    6
      IF ((0.0.GT.C).OR.(C.GT.1.0)) PRINT 3, ID,C,IDNUM                     7
      RETURN                                                                8
C                                                                           9
C                                                                          10
C                                                                          11
1     FORMAT (10X,*ITEM =*,I4,* HAS THE INVALID A PARAMETER OF *,F5.2,*    12
     1  A MUST BE GREATER THAN 0.0*,/10X,*ERROR FOUND *,*FOR THE SUBJECT    13
     2 WITH ID =*,A9)                                                      14
2     FORMAT (10X,*ITEM =*,I4,* HAS THE EXTREME B PARAMETER OF *,F5.2,5X    15
     1,*ERROR FOUND FOR THE SUBJECT WITH ID =*,A9)                         16
3     FORMAT (10X,*ITEM =*,I4,* HAS THE INVALID C PARAMETER OF *,F5.2,*    17
     1  C MUST BE BETWEEN 0.0 AND 1.0*,/10X,*ERROR FOUND *,*FOR THE SUBJ    18
     2ECT WITH ID =*,A9)                                                   19
      END                                                                  20
      SUBROUTINE BAYES (M,ITM,RESP,N,A,B,C,BTHET,BVAR,TINFO,EXPTOT)         1
      INTEGER RESP(M),ITM(M)                                                2
      REAL A(N),B(N),C(N)                                                   3
      DO 1 I=1,M                                                            4
      IF (ITM(I).EQ.0) GO TO 1                                              5
      CALL BSCOR (BTHET,BVAR,B(I),A(I),C(I),RESP(I))                        6
1     CONTINUE                                                              7
      CALL NOSTAT (M,ITM,A,B,C,BTHET,TINFO,EXPTOT)                          8
      RETURN                                                                9
      END                                                                  10
      SUBROUTINE BSCOR (BTHET,BVAR,DIF,DIS,GUESP,IRESP)                     1
      D=(DIF-BTHET)/SQRT(2.0*(1.0/DIS**2+BVAR))                             2
      ERFD=ERFNP(D)                                                         3
      EDSQ=EXP(D**2)                                                        4
      IF (EDSQ.EQ.0.0) RETURN                                               5
      EDSQ1=1.0/EDSQ                                                        6
      XKINV=0.5*(1.0-ERFD)                                                  7
      XLINV=GUESP+(1.0-GUESP)*XKINV                                         8
      IF ((XLINV.EQ.0.0).OR.(XKINV.EQ.0.0)) RETURN                          9
      XL=1.0/XLINV                                                         10
      IF (IRESP.NE.1) GO TO 1                                              11
      S=0.398942*(SQRT(BVAR)/SQRT(1.0+(1.0/DIS**2)/BVAR))*(1.0/XKINV)*ED   12
     1SQI                                                                  13
      T=1.0-1./72454*D*EDSQ*(1.0-ERFD)                                     14
      BTHET=BTHET+(1.0-GUESP)*XKINV*XL*S                                   15
      BVAR=BVAR-(1.0-GUESP)*XKINV*XL*S**2*(T-GUESP*XL)                     16
```

```
      RETURN                                                            17
1     BTHET=BTHET-0.797885*(BVAR/SQRT(1.0/DIS**2+BVAR))*EDSQI*(1.0/(1.0+ 18
     1ERFD))                                                            19
      PART1=1.128379/(1.0+(1.0/DIS**2)*(1.0/BVAR))                      20
      PART2=1.0/(EDSQ*(1.0+ERFD))**2                                    21
      PART3=0.564190+D*EDSQ*(1.0+ERFD)                                  22
      BVAR=BVAR*(1.0-PART1*PART2*PART3)                                 23
      RETURN                                                            24
      END                                                               25
      REAL FUNCTION ERFNP (X)                                            1
      DATA A1/0.254830/                                                  2
      DATA A2/-0.284497/                                                 3
      DATA A3/1.421414/                                                  4
      DATA A4/-1.453152/                                                 5
      DATA A5/1.061405/                                                  6
      DATA P/0.327591/                                                   7
      ERFNP=0.0                                                          8
      IF (X.EQ.0.0) RETURN                                               9
      ES=SIGN(1.0,X)                                                    10
      Y=ABS(X)                                                          11
      IF (Y.LT.6.0) GO TO 1                                             12
      ERFNP=ES                                                          13
      RETURN                                                            14
1     Y2=Y*Y                                                            15
      T=1.0/(1.0+P*Y)                                                   16
      AT=((A1+(A2+(A3+(A4+A5*T)*T)*T)*T)*T)                             17
      EAT=AT/EXP(Y2)                                                    18
      ERFNP=(1.0-EAT)*ES                                                19
      RETURN                                                            20
      END                                                               21
      SUBROUTINE MAXLK (M,ITM,RESP,N,A,B,C,MAX,EPS,IFAIL,SDRV,THETA,TINF  1
     10,EXPTOT,NUMITS,SEM)                                              2
      EXTERNAL FDDLOG,SDDLOG                                             3
      INTEGER RESP(M)                                                    4
      DIMENSION A(N), B(N), C(N), ITM(M)                                 5
C***  USES MAXIMUM LIKELIHOOD LOGISTIC SCORING ALGORITHM AND RESPONSE    6
C***  MODEL                                                              7
C***  BISECTION IS USED TO PROVIDE THE INITIAL GUESS FOR THE             8
C***  NEWTON-RAPHSON METHOD                                              9
      CALL BISECT (FDDLOG,RESP,A,B,C,M,ITM,5,GUESS)                     10
C                                                                       11
      CALL NEWTRAP (FDDLOG,SDDLOG,RESP,A,B,C,M,ITM,MAX,EPS,NUMITS,GUESS, 12
     1THETA,SDRV,IFAIL)                                                 13
C                                                                       14
      IF (IFAIL.EQ.1) 1,2                                               15
C***  NEWTON RAPHSON DID NOT CONVERGE                                   16
1     CALL NWTERR (THETA,SDRV,SEM,TINFO,EXPTOT)                         17
      RETURN                                                            18
C                                                                       19
2     CALL LGSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                      20
      SEM=1.0/SQRT(ABS(SDRV))                                           21
      RETURN                                                            22
      END                                                               23
      FUNCTION FDDLOG (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA XMAX,XMIN/200.0,-200.0/                                       4
      DATA D,RIGHT/1.7,1/                                                5
C***  CALCULATES FIRST DERIVATIVE OF LOG-LIKELIHOOD FUNCTION OF A        6
C***  RESPONSE VECTOR FOR THE LOGISTIC MODEL                             7
      SUM=0.0                                                            8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0) GO TO 1                                          10
      X=D*A(I)*(THETA-B(I))                                             11
      IF (X.LT.XMIN) X=XMIN                                             12
      IF (X.GT.XMAX) X=XMAX                                             13
```

```
      EXP=EXP(X)                                                        14
      AE=A(I)*EXP                                                       15
      SUM=SUM-AE/(EXP+1.0)                                              16
      IF (RESP(I).NE.RIGHT) GO TO 1                                     17
      CE=C(I)+EXP                                                       18
      SUM=SUM+AE/CE                                                     19
1     CONTINUE                                                          20
      FDDLOG=-1.7*SUM                                                   21
      RETURN                                                            22
      END                                                               23
      FUNCTION SDDLOG (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION ITM(M), A(M), B(M), C(M)                                 3
      DATA XMAX,XMIN/200.0,-200.0/                                       4
      DATA D,RIGHT/1.7,1/                                                5
C***  CALCULATES SECOND DERIVATIVE OF LOG-LIKELIHOOD FUNCTION            6
C***  OF A RESPONSE VECTOR FOR THE LOGISTIC MODEL                        7
      SUM=0.0                                                            8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0) GO TO 1                                          10
      X=D*A(I)*(THETA-B(I))                                             11
      IF (X.LT.XMIN) X=XMIN                                             12
      IF (X.GT.XMAX) X=XMAX                                             13
      EXP=EXP(X)                                                        14
      AE=A(I)*EXP                                                       15
      SUM=SUM-A(I)*AE/((1.0+EXP)*(1.0+EXP))                             16
      IF (RESP(I).NE.RIGHT) GO TO 1                                     17
      CE=C(I)+EXP                                                       18
      SUM=SUM+A(I)*C(I)*AE/(CE*CE)                                      19
1     CONTINUE                                                          20
      SDDLOG=-2.89*SUM                                                  21
      RETURN                                                            22
      END                                                               23
      SUBROUTINE BISECT (F1,RESP,A,B,C,M,ITM,NITER,BMID)                 1
      INTEGER RESP(M)                                                    2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
C***  CALCULATES APPROXIMATE ROOT OF F1 BY BISECTION:                    4
C***  BISECTING NITER (NUMBER OF ITERATIONS) TIMES.                      5
C***  BMID IS BEST CURRENT GUESS AT ROOT THETA                           6
C                                                                        7
C***  INITIALIZE LEFT BOUND AND F1(BOUND) AND RIGHT BOUND F1(BOUND)      8
      BL=-5.0                                                            9
      BR=5.0                                                            10
      BMID=0.0                                                          11
      TL=F1(RESP,ITM,A,B,C,M,BL)                                        12
      TR=F1(RESP,ITM,A,B,C,M,BR)                                        13
C***  TEST FOR NO ROOT IN INTERVAL--RETURN IF NO  SOLUTION              14
      IF ((TL*TR).GT.0.0) RETURN                                        15
C                                                                       16
C***  NOW CALCULATE BISECTIONS NITER TIMES                              17
      DO 3 I=1,NITER                                                    18
      TMID=F1(RESP,ITM,A,B,C,M,BMID)                                    19
      IF ((TMID*TL).GT.0.0) GO TO 1                                     20
C***  REPLACE RIGHT BOUND WITH BMID                                     21
      BR=BMID                                                           22
      GO TO 2                                                           23
C***  REPLACE LEFT BOUND WITH BMID                                      24
1     TL=TMID                                                           25
      BL=BMID                                                           26
C***  FIND NEW MIDPOINT BMID                                            27
2     BMID=(BL+BR)/2.0                                                  28
3     CONTINUE                                                          29
      RETURN                                                            30
      END                                                               31
      SUBROUTINE NEWTRAP (F1,F2,RESP,A,B,C,M,ITM,NITER,EPS,NUMITS,GUESS,  1
     1THETA,SDRV,IFAIL)                                                 2
```

```
      INTEGER RESP(M)                                                    3
      DIMENSION A(M), B(M), C(M), ITM(M)                                 4
C***  CALCULATES THE ROOT OF F1 GIVEN ITS FIRST DERIVATIVE F2            5
C***  AND AN INITIAL GUESS USING NEWTON-RAPHSON METHOD                   6
C***  THETA IS APPR. TO THE ROOT; SDRV IS F2(THETA)                      7
      NUMITS=0                                                           8
      THETA=GUESS                                                        9
C***  LOOP UNTIL ERR<EPS OR NUMBER OF ITERATIONS BECOMES TOO LARGE       10
1     FDRV=F1(RESP,ITM,A,B,C,M,THETA)                                    11
      SDRV=F2(RESP,ITM,A,B,C,M,THETA)                                    12
      ERR=FDRV/SDRV                                                      13
      THETA=THETA-ERR                                                    14
      NUMITS=NUMITS+1                                                    15
C***  EXIT LOOP CRITERION                                                16
      IF ((NUMITS.LT.NITER).AND.(ABS(ERR).GT.EPS)) GO TO 1               17
C***  END LOOP.  TEST FOR CONVERGENCE AND SET IFAIL                      18
      IFAIL=0                                                            19
      IF (ABS(ERR).LT.EPS) RETURN                                        20
C                                                                        21
C***  NEWTON RAPHSON METHOD DOES NOT CONVERGE                            22
      IFAIL=1                                                            23
      RETURN                                                             24
      END                                                                25
      SUBROUTINE NWTERR (THETA,SFORM,SEM,TINFO,EXPTOT)                    1
C***  SETS ERROR VALUES FOR THE CASE IN WHICH NEWTON RAPHSON FAILS       2
C***  TO CONVERGE                                                        3
      THETA=-99.99                                                       4
      SFORM=-99.99                                                       5
      SEM=-99.99                                                         6
      TINFO=-99.99                                                       7
      EXPTOT=-99.99                                                      8
      RETURN                                                             9
      END                                                                10
      SUBROUTINE LGSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                  1
      DIMENSION A(M), B(M), C(M), ITM(M)                                 2
      DATA XMAX,XMIN/12.0,-12.0/                                         3
      TINFO=0.0                                                          4
      EXPTOT=0.0                                                         5
      KOUNT=0                                                            6
C                                                                        7
      DO 1 I=1,M                                                         8
      IF (ITM(I).EQ.0) GO TO 1                                           9
      KOUNT=KOUNT+1                                                      10
      ARGU=-1.7*A(I)*(THETA-B(I))                                        11
      IF (ARGU.GT.XMAX) ARGU=XMAX                                        12
      IF (ARGU.LT.XMIN) ARGU=XMIN                                        13
      P=C(I)+(1.0-C(I))*(1.0/(1.0+EXP(ARGU)))                            14
      Q=1.0-P                                                            15
      EARG=EXP(-ARGU)                                                    16
      PPRIME=EARG/((1.0+EARG)*(1.0+EARG))                                17
      PPRIME=PPRIME*(1.0-C(I))*A(I)*1.7                                  18
      TINFO=TINFO+(PPRIME*PPRIME)/(P*Q)                                  19
      EXPTOT=EXPTOT+P                                                    20
1     CONTINUE                                                           21
      EXPTOT=EXPTOT/KOUNT                                                22
      RETURN                                                             23
      END                                                                24
      SUBROUTINE MAXLNO (RESP,ITM,M,N,A,B,C,MAX,EPS,THETA,SDRV,IFAIL,TIN  1
     1FO,EXPTOT,NUMITS,SEM)                                              2
      EXTERNAL FDNOGV,SDNOGV                                             3
      INTEGER RESP(M)                                                    4
      DIMENSION ITM(N), A(N), B(N), C(N)                                 5
C***  USES MAXIMUM LIKELIHOOD NORMAL OGIVE SCORING ALGORITHM AND         6
C***  RESPONSE VECTOR                                                    7
C***  BISECTION IS USED TO PROVIDE THE INITIAL GUESS FOR THE             8
C***  NEWTON RAPHSON METHOD                                              9
```

```
C                                                                       10
      CALL BISECT (FDNOGV,RESP,A,B,C,M,ITM,5,GUESS)                     11
C                                                                       12
      CALL NEWTRAP (FDNOGV,SDNOGV,RESP,A,B,C,M,ITM,MAX,EPS,NUMITS,GUESS,13
     1THETA,SDRV,IFAIL)                                                 14
      IF (IFAIL.EQ.1) 1,2                                               15
C***  NEWTON RAPHSON DID NOT CONVERGE                                   16
1     CALL NWTERR (THETA,SDRV,SEM,TINFO,EXPTOT)                         17
      RETURN                                                            18
C                                                                       19
2     CALL NOSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                      20
      SDRV=ABS(SDRV)                                                    21
      SEM=1.0/SQRT(SDRV)                                                22
      RETURN                                                            23
      END                                                               24
      FUNCTION FDNOGV (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA PI,RIGHT/3.141592,1/                                          4
      DATA XMAX,XMIN/7.0,-7.0/                                           5
C***  CALCULATES FIRST DERIVATIVE OF LOG-LIKELIHOOD FUNCTION OF          6
C***  A RESPONSE VECTOR FOR THE NORMAL OGIVE MODEL                       7
C                                                                        8
      SUM=0.0                                                            9
      ROOTPI=1.0/SQRT(2.0*PI)                                           10
      DO 2 I=1,M                                                        11
      IF (ITM(I).EQ.0) GO TO 2                                          12
      TEMP=A(I)*(THETA-B(I))                                            13
      IF (TEMP.GT.XMAX) TEMP=XMAX                                       14
      IF (TEMP.LT.XMIN) TEMP=XMIN                                       15
      X=-(TEMP*TEMP)/2.0                                                16
      DNMRAT=ROOTPI*A(I)*(1.0-C(I))*EXP(X)                              17
      DENOM=C(I)+(1.0-C(I))*CDFN(TEMP)                                  18
      IF (RESP(I).EQ.RIGHT) GO TO 1                                     19
      DENOM=-(1.0-DENOM)                                                20
1     SUM=SUM+(DNMRAT/DENOM)                                            21
2     CONTINUE                                                          22
      FDNOGV=SUM                                                        23
      RETURN                                                            24
      END                                                               25
      FUNCTION SDNOGV (RESP,ITM,A,B,C,M,THETA)                           1
      INTEGER RESP(M),RIGHT                                              2
      DIMENSION A(M), B(M), C(M), ITM(M)                                 3
      DATA PI,RIGHT/3.141592,1/                                          4
      DATA XMAX,XMIN/7.0,-7.0/                                           5
C***  CALCULATES SECOND DERIVATIVE OF LOG-LIKELIHOOD FUNCTION            6
C***  OF A RESPONSE VECTOR FOR THE NORMAL OGIVE MODEL                    7
C                                                                        8
      SUM=0.0                                                            9
      ROOTPI=1.0/SQRT(2.0*PI)                                           10
      DO 2 I=1,M                                                        11
      IF (ITM(I).EQ.0) GO TO 2                                          12
      TEMP1=A(I)*(THETA-B(I))                                           13
      IF (TEMP1.GT.XMAX) TEMP1=XMAX                                     14
      IF (TEMP1.LT.XMIN) TEMP1=XMIN                                     15
      X=-TEMP1*TEMP1/2.0                                                16
      TEMP2=ROOTPI*(1.0-C(I))*A(I)*EXP(X)                               17
      FIRNUM=TEMP2*TEMP2                                                18
      SECNUM=TEMP2*A(I)*TEMP1                                           19
      SDENOM=C(I)+(1.0-C(I))*CDFN(TEMP1)                                20
      FDENOM=SDENOM*SDENOM                                              21
      IF (RESP(I).EQ.RIGHT) GO TO 1                                     22
      FDENOM=(1.0-SDENOM)*(1.0-SDENOM)                                  23
      SDENOM=-(1.0-SDENOM)                                              24
1     SUM=SUM-(FIRNUM/FDENOM)-(SECNUM/SDENOM)                           25
2     CONTINUE                                                          26
```

```
      SONOGV=SUM                                                        27
      RETURN                                                            28
      END                                                               29
      SUBROUTINE NOSTAT (M,ITM,A,B,C,THETA,TINFO,EXPTOT)                 1
      DIMENSION A(M), B(M), C(M), ITM(M)                                 2
      DATA XMAX,XMIN/7.0,-7.0/                                           3
      DATA PI/3.141592/                                                  4
      TINFO=0.0                                                          5
      EXPTOT=0.0                                                         6
      KOUNT=0                                                            7
C                                                                        8
      DO 1 I=1,M                                                         9
      IF (ITM(I).EQ.0.0) GO TO 1                                        10
      KOUNT=KOUNT+1                                                     11
      TEMP=A(I)*(THETA-B(I))                                            12
      IF (TEMP.GT.XMAX) TEMP=XMAX                                       13
      IF (TEMP.LT.XMIN) TEMP=XMIN                                       14
      P=C(I)+(1.0-C(I))*CDFN(TEMP)                                      15
      Q=1.0-P                                                           16
      TEMP=-TEMP*TEMP/2.0                                               17
      PPRIME=(1.0/SQRT(2.0*PI))*(1.0-C(I))*A(I)*EXP(TEMP)               18
      TINFO=TINFO+(PPRIME*PPRIME)/(P*Q)                                 19
      EXPTOT=EXPTOT+P                                                   20
1     CONTINUE                                                          21
      EXPTOT=EXPTOT/KOUNT                                               22
      RETURN                                                            23
      END                                                               24
```

APPENDIX E
LINPSCO FORTRAN PROGRAM LISTING

```
      PROGRAM LINPSCO (INPUT,OUTPUT,DATA,IPOOL,TAPE1=DATA,TAPE2=IPOOL,TA
     1PE3,PUNCH)
C*****************************************************************
C              DEFINITION OF VARIABLES USED IN LINPSCO
C
C     T=THETA:  THE ABILITY ESTIMATE
C     ITEM:  AN ARRAY OF THE ITEMS IN THE ITEM POOL
C     INUP:  THE NUMBER OF ITEMS IN THE ITEM POOL
C     A:  AN ARRAY OF THE DISCRIMINATION PARAMETER(S) ASSOCIATED WITH
C       EACH ITEM
C   B:  AN ARRAY OF THE DIFFICULTY PARAMETERS ASSOCIATED WITH EACH ITEM
C     MAXCAT:  NO. OF RESPONSE CATEGORIES MINUS 1 FOR THE ITEM IN THE
C           ITEM POOL WHICH HAS THE MAXIMUM NO. OF RESPONSE
C           CATEGORIES
C     INAD:  AN ARRAY OF THE ITEMS WHICH HAVE BEEN ADMINISTERED TO ALL
C          SUBJECTS
C     M:  THE NUMBER OF ITEMS IN THE TEST
C     ADM AND BDM:  THE ARRAYS OF PARAMETERS ASSOCIATED WITH THE ITEMS
C                 IN INAD
C     INADS:  AN ARRAY OF THE ITEMS USED TO CALCULATE THETA FOR A GIVEN
C           SUBJECT (INADS=INAD-(REJECTION AND OMISSIONS) WHERE ID S
C           OF REJECTED OR OMITTED ITEMS ARE SET TO ZERO)
C     IRESP:  THE RESPONSE VECTOR OF THE CURRENT SUBJECT
C     NCAT:  AN ARRAY OF THE NUMBER OF RESPONSE CATEGORIES MINUS 1 FOR
C          EACH ITEM ADMINISTERED
C     PERCNT:  PERCENTAGE OF ITEMS ANSWERED BY A SUBJECT FOR WHICH THE
C           BEST RESPONSE HAS BEEN CHOSEN
C     D:  A CONSTANT USED TO CHANGE THE METRIC OF THE LOGISTIC GRADED
C       MODEL.  D=1.7 OR 1.0; THE DEFAULT VALUE IS 1.0
C
C*****************************************************************
      COMMON D
      DIMENSION ITEM(100), A(100,10), B(100,10), INAD(100), IREJ(100), I
     1RESP(100), ADM(100,10), BDM(100,10), NAME(2), IFORM(8), NCAT(100),
     2 IFORM1(8), INADS(100), DESC(24)
      INTEGER OPT1,OPT2,OPT4,ADIM,BDIM
      N=NC=0
C*****************************************************************
C                                                               *
CREAD OPTIONS AND PROGRAM PARAMETER FROM INPUT FILE DATA IS ON TAPE2 *
C                                                               *
C*****************************************************************
      IPOOL=2
      READ 30, INUP,M,MAXCAT,OPT1,OPT2,OPT4,D,IOMIT
      CALL CHKINP (INUP,M,OPT4)
      IF (D.NE.1.7) D=1.0
      READ 33, (IFORM1(I),I=1,8)
      READ 32, (NCAT(I),I=1,M)
C***  INPUT ITEMS AND THEIR PARAMETERS
      IKI=ADIM(MAXCAT,OPT4)
      K=BDIM(MAXCAT,OPT4)
      DO 1 I=1,INUP
1     READ (IPOOL,IFORM1) ITEM(I),(A(I,JJ),JJ=1,IKI),(B(I,J),J=1,K)
C
      READ 31, (INAD(I),I=1,M)
      READ 32, (IREJ(I),I=1,M)
      READ 33, (IFORM(I),I=1,8)
      READ 34, DESC
      PRINT 35, INUP,M,IOMIT,OPT1,OPT2,OPT4,MAXCAT,D,IFORM1,IFORM,DESC
C
C***  THOSE ITEM ID S WHICH ARE IN THE REJECTION VECTOR WILL BE SET
C***  TO ZERO.  ZERO ITEM ID S WILL BE SKIPPED DURING COMPUTATIONS
      DO 2 I=1,M
2     IF (IREJ(I).EQ.1) INAD(I)=0
      IF (OPT2.EQ.1) GO TO 10
C
```

```
C                                                                       67
C***  SEARCH FOR ITEMS ADMINISTERED IN POOL                             68
      DO 9 J=1,M                                                        69
      IKI=ADIM(NCAT(J),OPT4)                                            70
      K=BDIM(NCAT(J),OPT4)                                              71
      IF (INAD(J).NE.0) GO TO 5                                         72
C***  BLANK OR REJECTED ITEM ENCOUNTERED                                73
      DO 3 L=1,IKI                                                      74
3     ADM(J,L)=0.0                                                      75
      DO 4 L=1,K                                                        76
4     BDM(J,L)=0.0                                                      77
      GO TO 9                                                           78
5     CONTINUE                                                          79
C                                                                       80
      DO 8 I=1,INUP                                                     81
      IF (INAD(J).NE.ITEM(I)) GO TO 8                                   82
      DO 6 L=1,IKI                                                      83
6     ADM(J,L)=A(I,L)                                                   84
      DO 7 L=1,K                                                        85
7     BDM(J,L)=B(I,L)                                                   86
      GO TO 9                                                           87
8     CONTINUE                                                          88
      INAD(J)=0                                                         89
9     CONTINUE                                                          90
      GO TO 13                                                          91
C                                                                       92
C***  ALL ITEMS IN POOL HAVE BEEN ADMINISTERED SO POOL DOES NOT         93
C***  NEED TO BE SEARCHED                                               94
10    DO 12 I=1,M                                                       95
      IKI=ADIM(NCAT(I),OPT4)                                            96
      K=BDIM(NCAT(I),OPT4)                                              97
      DO 11 JJ=1,IKI                                                    98
11    ADM(1,JJ)=A(I,JJ)                                                 99
      DO 12 J=1,K                                                      100
      BDM(I,J)=B(I,J)                                                  101
12    CONTINUE                                                         102
C                                                                      103
C                                                                      104
C***  PUNCH THE ITEM PARAMETERS EXTIMATES CORRESPONDING TO THE ITEMS   105
C***  IN THE TEST                                                      106
13    IF (OPT1.NE.1) GO TO 15                                          107
      DO 14 I=1,M                                                      108
      IF (INAD(I).EQ.0) GO TO 14                                       109
      IKI=ADIM(NCAT(I),OPT4)                                           110
      K=BDIM(NCAT(I),OPT4)                                             111
      PUNCH 36, INAD(I),IREJ(I),(ADM(I,JJ),JJ=1,IKI)                   112
      PUNCH 37, (BDM(I,J),J=1,K)                                       113
14    CONTINUE                                                         114
C                                                                      115
C     PRINT OUT THE ITEMS AND THEIR PARAMETERS                         116
15    DO 16 III=1,M                                                    117
      IKI=ADIM(NCAT(III),OPT4)                                         118
      K=BDIM(NCAT(III),OPT4)                                           119
      PRINT 39, INAD(III),IREJ(III),(ADM(III,JJ),JJ=1,IKI)             120
16    PRINT 40, (BDM(III,J),J=1,K)                                     121
      PRINT 43                                                         122
C**********************************************************************123
C***                                                                  *124
C***  READ SUBJECT FROM TAPE1, CALCULATE THETA, LOOP BACK TO 300 UNTIL*125
C***  THERE ARE NO MORE SUBJECTS TO SCORE                             *126
C***                                                                  *127
C**********************************************************************128
      N=0                                                              129
17    READ (1,IFORM) NAME,ID,(IRESP(I),I=1,M)                          130
      IO=UNIT(1)                                                       131
      IF (IO.LE.0) GO TO 18                                            132
```

```
      PRINT 38, ID                                                          133
18    IF (ID.EQ.0) GO TO 29                                                 134
C                                                                           135
C***  CHECK FOR VALID RESPONSES AND FOR SPECIAL RESPONSE VECTORS            136
      CALL CHKRSP (IRESP,IOMIT,INAD,NCAT,M,NAME,ID,INADS)                   137
C                                                                           138
      CALL CHKVEC (IRESP,INADS,NCAT,M,ICHK,NBEST,NANS)                      139
      GO TO (19,20,21,23), ICHK                                             140
C***  ALL ITEMS OMITTED                                                     141
19    PRINT 45, NAME,ID                                                     142
      T=-50.00                                                              143
      PERCNT=-50.00                                                         144
      GO TO 22                                                              145
C***  ALL RESPONSES INCORRECT                                               146
20    PRINT 46, NAME,ID                                                     147
      T=-10.0                                                               148
      PERCNT=0.0                                                            149
      GO TO 22                                                              150
C***  ALL RESPONSES CORRECT                                                 151
21    PRINT 47, NAME,ID                                                     152
      T=10.0                                                                153
      PERCNT=1.0                                                            154
C                                                                           155
22    SEORM=0.0                                                             156
      NUMITS=0                                                              157
      TINFO=0.0                                                             158
      GO TO 28                                                              159
23    CONTINUE                                                              160
C                                                                           161
      PERCNT=FLOAT(NBEST)/FLOAT(NANS)                                       162
C                                                                           163
C                                                                           164
C***  DETERMINE WHICH MODEL TO USE                                          165
      IF (OPT4-2) 24,25,26                                                  166
24    CALL LOGRAD (IRESP,ADM,BDM,M,NCAT,INADS,.001,50,SEORM,IFAIL,T,NUMI    167
     1TS,TINFO,SE)                                                          168
      GO TO 27                                                              169
25    CALL NOGRAD (IRESP,ADM,BDM,M,NCAT,INADS,.001,50,SEORM,IFAIL,T,NUMI    170
     1TS,TINFO,SE)                                                          171
      GO TO 27                                                              172
26    CALL NOMLOG (IRESP,ADM,BDM,M,NCAT,INADS,.001,50,SEORM,IFAIL,T,NUMI    173
     1TS,TINFO,SE)                                                          174
C                                                                           175
C***  OUTPUT RESULTS TO TAPE3                                               176
27    IF (IFAIL.EQ.0) GO TO 28                                              177
C***  CONVERGENCE NOT OBTAINED                                              178
      PRINT 41, NAME,ID                                                     179
      NC=NC+1                                                               180
C                                                                           181
28    WRITE (3,42) NAME,ID,PERCNT,T,SEORM,NANS,NUMITS,TINFO,SE              182
      N=N+1                                                                 183
      GO TO 17                                                              184
C                                                                           185
29    PRINT 44, N,NC                                                        186
      STOP                                                                  187
C                                                                           188
30    FORMAT (2I4,I1,2I1,1X,I1,F5.2,27X,I2)                                 189
31    FORMAT (16I5)                                                         190
32    FORMAT (80I1)                                                         191
33    FORMAT (8A10)                                                         192
34    FORMAT (8A10)                                                         193
35    FORMAT (T50,*LINPSCO*,/,T50*======*,//////,T20,*LINEAR POLYCHOTOMU    194
     1S SCORING WITH TWO PARAMETER MODELS*,////,T40*PSYCHOMETRICS METHOD    195
     2S PROGRAM*,/,T40*DEPARTMENT OF PSYCHOLOGY*,/,T40*UNIVERSITY OF MIN    196
     3NESOTA*,/,T40*MPLS.  MINN. 55455*,////,T20,*INUP*,T27,*=*,I4,/,T20    197
     4,*MMAX*,T27,*=*I4./.T20.*IOMIT*.T27,*=*.I4./.T20*OPT1*,T27*=*I4./.    198
```

```
     >I2,*OPT2*,T27*=*,I4,/,T20*OPT4*,T27*=*,I4,/,T20,*MAXCAT*,T27,*=*,I      199
     64,/T20,*D*,T27,*=*,F4.1,/T20,*VARIABLE FORMAT*,* FOR POOL =*,8A10,      200
     7/T20,*VARIABLE FORMAT FOR DATA =*,8A10,/T20,8A10,/T20,8A10,/,T20,8      201
     8A10)                                                                    202
36    FORMAT (I5,1X,I2,2X,10F6.2)                                             203
37    FORMAT (10F6.2)                                                         204
38    FORMAT (*PARITY ERROR ON TAPE*,100X,I2)                                 205
39    FORMAT (/10X,*ITEM ID = *,I5,5X,*REJECTION =*,I3,/12X,*A:  *,10F6.      206
     12)                                                                      207
40    FORMAT (12X,*B:  *,10F6.2)                                              208
41    FORMAT (10X,*CONVERGENCE NOT OBTAINED FOR SUBJECT =*,2A10,* ID =*,      209
     1A9)                                                                     210
42    FORMAT (1X,2A10,A9,F5.2,2F7.2,I4,I4,2F7.2)                              211
43    FORMAT (////)                                                           212
44    FORMAT (10X,*CASES READ =*,I5,*  CASES NOT CONVERGED =*,I5)             213
45    FORMAT (10X,*ALL ITEMS OMITTED FOR SUBJECT =*,2A10,* ID =*,A9)          214
46    FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS ALL ANSWERS IN*,*COR      215
     1RECT*)                                                                  216
47    FORMAT (10X,*SUBJECT =*,2A10,* ID =*,A9,* HAS ALL ANSWERS COR*,*RE      217
     1CT*)                                                                    218
      END                                                                     219
      INTEGER FUNCTION ADIM (NUMCAT,OPT4)                                       1
      INTEGER OPT4                                                              2
C***  DETERMINES THE NUMBER OF A PARAMETERS FOR A GIVEN ITEM                    3
      ADIM=1                                                                    4
      IF (OPT4.EQ.3) ADIM=NUMCAT+1                                              5
      RETURN                                                                    6
      END                                                                       7
      INTEGER FUNCTION BDIM (NUMCAT,OPT4)                                       1
      INTEGER OPT4                                                              2
C***  DETERMINES THE NUMBER OF B PARAMETERS FOR A GIVEN ITEM                    3
      BDIM=NUMCAT                                                               4
      IF (OPT4.EQ.3) BDIM=NUMCAT+1                                              5
      RETURN                                                                    6
      END                                                                       7
      SUBROUTINE CHKINP (INUP,M,OPT4)                                           1
      INTEGER OPT4                                                              2
C***  CHECKS FOR ERRORS IN THE INPUT DATA                                       3
C***  IF AN ERROR IS FOUND, A MESSAGE IS PRINTED AND THE PROGRAM                4
C***  HALTS                                                                     5
C                                                                               6
      IERR=0                                                                    7
      IF (INUP.LE.100) GO TO 1                                                  8
      PRINT 4, INUP                                                             9
      IERR=1                                                                   10
1     IF (M.LE.INUP) GO TO 2                                                   11
      PRINT 5, M,INUP                                                          12
      IERR=1                                                                   13
2     IF (1.LE.OPT4.AND.OPT4.LE.3) GO TO 3                                     14
      PRINT 6, OPT4                                                            15
      IERR=1                                                                   16
3     IF (IERR.EQ.1) STOP                                                      17
      RETURN                                                                   18
C                                                                              19
C                                                                              20
4     FORMAT (10X,*INPUT ERROR:  NO. OF ITEMS IN ITEM POOL =*,I5,/10X,*N       21
     1O. MUST BE .LE. 100*)                                                    22
5     FORMAT (10X,*INPUT ERROR:  NO. ITEMS ADMINISTERED =*,I5,/10X,*NO.        23
     1MUST BE .LE. NO. OF ITEMS IN ITEM POOL =*,I5)                            24
6     FORMAT (10X,*INPUT ERROR:  OPTION 4 =*,I3,* DOES NOT CORRES*,*POND       25
     1 TO ANY OF THE AVAILABLE RESPONSE MODELS*)                               26
      END                                                                      27
      SUBROUTINE NOGRAD (IRESP,A,B,NITEMS,NCAT,INAD,EPS,MAXIT,SFORM,IFAI        1
     1L,THETA,NOMITS,TINFO,SE)                                                 2
      EXTERNAL FDEVNO,SDEVNO                                                    3
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100)          4
```

```
      CALL BISECT (FORVNO,IRESP,A,B,NITEMS,NCAT,INAD,5,GUESS)                5
      CALL NEWTRAP (FORVNO,SDRVNO,IRESP,A,B,NITEMS,NCAT,INAD,MAXIT,EPS,N     6
     1UMITS,GUESS,THETA,SDRV,IFAIL)                                          7
      SEORM=-SDRV                                                            8
      IF (IFAIL.EQ.1) GO TO 1                                                9
      CALL NOINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)                         10
      SE=1.0/SQRT(ABS(SDRV))                                                 11
      RETURN                                                                 12
C                                                                            13
1     TINFO=-99.99                                                           14
      SE=-99.99                                                              15
      RETURN                                                                 16
      END                                                                    17
      FUNCTION IVALUE (IRESP,NCAT)                                            1
C***  CHECKS RESPONSE FOR SPECIAL CASES                                       2
C***  IVALUE RETURNS . . . 1  IF IRESP IS BEST RESPONSE                       3
C***                       2  IF IRESP IS WORST RESPONSE                     4
C***                       3  OTHERWISE                                      5
C                                                                             6
      IVALUE=3                                                                7
      IF (IRESP.EQ.1) IVALUE=1                                                8
      IF (IRESP.EQ.(NCAT+1)) IVALUE=2                                         9
      RETURN                                                                 10
      END                                                                    11
      SUBROUTINE RCAT1 (A,B,THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)                1
C***  COMPUTES VALUES NECESSARY FOR THE CALCULATION OF THE DERIV-             2
C***  ATIVES OF THE NORMAL OGIVE GRADED MODEL FOR THE SPECIAL CASE            3
C***  WHEN IRESP IS THE BEST RESPONSE                                         4
C                                                                             5
      TMINB0=THETA-B                                                          6
      TMINB1=0.0                                                              7
      Y=A*TMINB0                                                              8
      P=CDFN(Y)                                                               9
      ETOZ0=EXP(-Y*Y/2.0)                                                    10
      ETOZ1=0.0                                                              11
      RETURN                                                                 12
      END                                                                    13
      SUBROUTINE RCATN (A,B,THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)                1
C***  COMPUTES VALUES NECESSARY FOR THE CALCULATION OF THE DER-               2
C***  IVATIVES OF THE NORMAL OGIVE GRADED MODEL FOR THE SPECIAL               3
C***  CASE WHEN IRESP IS THE WORST RESPONSE                                   4
C                                                                             5
      TMINB0=0.0                                                              6
      TMINB1=THETA-B                                                          7
      Y1=A*TMINB1                                                             8
      P=1.0-CDFN(Y1)                                                          9
      ETOZ0=0.0                                                              10
      ETOZ1=EXP(-Y1*Y1/2.0)                                                  11
      RETURN                                                                 12
      END                                                                    13
      SUBROUTINE RCATOT (A,B0,B1,THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)           1
C***  COMPUTES VALUES NECESARY FOR THE CALCULATION OF DERIVATIVES OF          2
C***  THE NORMAL OGIVE GRADED MODEL FOR ALL OTHER CASES                       3
C                                                                             4
      TMINB0=THETA-B0                                                         5
      TMINB1=THETA-B1                                                         6
      Y=A*TMINB0                                                              7
      Y1=A*TMINB1                                                             8
      P=CDFN(Y)-CDFN(Y1)                                                      9
      ETOZ0=EXP(-Y*Y/2.0)                                                    10
      ETOZ1=EXP(-Y1*Y1/2.0)                                                  11
      RETURN                                                                 12
      END                                                                    13
      FUNCTION FORVNO (IRESP,INAD,NITEMS,NCAT,A,B,THETA)                      1
      DIMENSION IRESP(100), INAD(100), NCAT(100), A(100,10), B(100,10)        2
C***  CALCULATES THE FIRST DERIVATIVE FOR THE NORMAL OGIVE GRADED MODEL       3
```

```
      SUM=0.0                                                                  4
      DO 5 I=1,NITEMS                                                          5
      IF (INAD(I).EQ.0) GO TO 5                                                6
      K=IRESP(I)                                                               7
      J=IVALUE(K,NCAT(I))                                                      8
      GO TO (1,2,3), J                                                         9
C                                                                             10
C***  IRESP IS BEST RESPONSE                                                  11
1     CALL PCAT1 (A(I,1),B(I,K),THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)            12
      GO TO 4                                                                 13
C                                                                             14
C***  IRESP IS WORST RESPONSE                                                 15
2     CALL PCATN (A(I,1),B(I,K-1),THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)          16
      GO TO 4                                                                 17
C                                                                             18
C***  ALL OTHER RESPONSES                                                     19
3     CALL PCATOT (A(I,1),B(I,K),B(I,K-1),THETA,TMINB0,TMINB1,P,ETOZ0,ET      20
     1OZ1)                                                                    21
4     CONTINUE                                                                22
C                                                                             23
      IF (P.EQ.0.0) P=0.0001                                                  24
      SUM=SUM+A(I,1)*(ETOZ0-ETOZ1)/P                                          25
5     CONTINUE                                                                26
C                                                                             27
      FORVNO=SUM/SQRT(2.0*3.142)                                              28
      RETURN                                                                  29
      END                                                                     30
      FUNCTION SDRVNO (IRESP,INAD,NITEMS,NCAT,A,B,THETA)                      31
      DIMENSION IRESP(100), INAD(100), NCAT(100), A(100,10), B(100,10)         1
C***  CALCULATES SECOND DERIVATIVE FOR THE NORMAL OGIVE GRADED MODEL           2
C                                                                              3
      SUM=0.0                                                                  4
      ROOTPI=1.0/SQRT(2.0*3.142)                                               5
      DO 5 I=1,NITEMS                                                          6
      IF (INAD(I).EQ.0) GO TO 5                                                7
      K=IRESP(I)                                                               8
      J=IVALUE(K,NCAT(I))                                                      9
      GO TO (1,2,3), J                                                        10
C                                                                             11
C***  IRESP IS BEST RESPONSE                                                  12
1     CALL PCAT1 (A(I,1),B(I,K),THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)            13
      GO TO 4                                                                 14
C                                                                             15
C***  IRESP IS WORST RESPONSE                                                 16
2     CALL PCATN (A(I,1),B(I,K-1),THETA,TMINB0,TMINB1,P,ETOZ0,ETOZ1)          17
      GO TO 4                                                                 18
C                                                                             19
C***  ALL OTHER RESPONSES                                                     20
3     CALL PCATOT (A(I,1),B(I,K),B(I,K-1),THETA,TMINB0,TMINB1,P,ETOZ0,ET      21
     1OZ1)                                                                    22
4     CONTINUE                                                                23
C                                                                             24
      IF (P.EQ.0.0) P=0.0001                                                  25
      SUM1=A(I,1)*ROOTPI*(ETOZ0-ETOZ1)/P                                      26
      SUM1=-SUM1*SUM1                                                         27
      SUM2=-(A(I,1)**3)*ROOTPI*((TMINB0*ETOZ0)-(TMINB1*ETOZ1))/P              28
      SUM=SUM+SUM1+SUM2                                                       29
5     CONTINUE                                                                30
C                                                                             31
      SDRVNO=SUM                                                              32
      RETURN                                                                  33
      END                                                                     34
      SUBROUTINE NCINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)                    35
      DIMENSION A(100,10), B(100,10), NCAT(100), INAD(100)                     1
C***  COMPUTES INFORMATION FOR GRADED NORMAL OGIVE MODEL AT GIVEN              2
```

```
C***  VALUE OF THETA                                                    4
C                                                                       5
      TINFO=0.0                                                         6
      ROOTPI=1.0/SQRT(2.0*3.142)                                        7
C***  LOOP OVER ITEMS                                                   8
      DO 5 I=1,NITEMS                                                   9
      IF (INAD(I).EQ.0) GO TO 5                                        10
C                                                                      11
C***  INITIALIZATION--VALUES CALCULATED FOR FIRST CATEGORY OF ITEM I   12
      Y=A(I,1)*(THETA-B(I,1))                                          13
      P0=CDFN(Y)                                                       14
      P1=0.0                                                           15
      ETOZ0=EXP(-Y*Y/2.0)                                              16
      ETOZ1=0.0                                                        17
      KATGRY=0                                                         18
C                                                                      19
C***  LOOP OVER ALL THE CATEGORIES OF ITEM I                           20
1     KATGRY=KATGRY+1                                                  21
      P=P0-P1                                                          22
      IF (P.EQ.0.0) P=0.0001                                           23
      FDRVP=A(I,1)*ROOTPI*(ETOZ0-ETOZ1)                                24
      TINFO=TINFO+(FDRVP*FDRVP)/P                                      25
C                                                                      26
      ETOZ1=ETOZ0                                                      27
      P1=P0                                                            28
C                                                                      29
      IF (KATGRY-(NCAT(I)+1)) 2,3,4                                    30
C***  CURRENT CATEGORY UNDER CONSIDERATION IS NOT ONE OF THE EXTREMES  31
2     Y=A(I,1)*(THETA-B(I,KATGRY))                                     32
      P0=CDFN(Y)                                                       33
      ETOZ0=EXP(-Y*Y/2.0)                                              34
      GO TO 1                                                          35
C***  LAST CATEGORY OF AN ITEM IS BEING CONSIDERED                     36
3     P0=1.0                                                           37
      ETOZ0=0.0                                                        38
      GO TO 1                                                          39
C***  ALL CATEGORIES FOR ITEM I HAVE BEEN EXAMINED                     40
4     CONTINUE                                                         41
C                                                                      42
5     CONTINUE                                                         43
      RETURN                                                           44
      END                                                              45
      SUBROUTINE LOGRAD (IRESP,A,B,NITEMS,NCAT,INAD,EPS,NITER,SFORM,IFAI  1
     1L,THETA,NUMITS,TINFO,SE)                                         2
      EXTERNAL FDRVLL,SDRVLL                                           3
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100) 4
      CALL BISECT (FDRVLL,IRESP,A,B,NITEMS,NCAT,INAD,5,GUESS)          5
      CALL NEWTRAP (FDRVLL,SDRVLL,IRESP,A,B,NITEMS,NCAT,INAD,NITER,EPS,N  6
     1UMITS,GUESS,THETA,SDRV,IFAIL)                                    7
      SFORM=-SDRV                                                      8
      IF (IFAIL.EQ.1) GO TO 1                                          9
      CALL LLINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)                   10
      SE=1.0/SQRT(ABS(SDRV))                                           11
      RETURN                                                           12
C                                                                      13
1     TINFO=-99.99                                                     14
      SE=-99.99                                                        15
      RETURN                                                           16
      END                                                              17
      SUBROUTINE CALCP (IRESP,INAD,NCAT,A,B,NITEMS,THETA,P)             1
      COMMON D                                                          2
      DIMENSION IRESP(100), INAD(100), NCAT(100), B(100,10), P(100,2), A  3
     1(100,10)                                                          4
C****  CALCULATES UPPER AND LOWER P FOR EACH ITEM WITH GIVEN ANSWER VECT  5
      DO 4 I=1,NITEMS                                                   6
      IF (INAD(I).EQ.0) GO TO 4                                         7
```

```
      J=IRESP(I)                                                          8
      IF (J.EQ.1) GO TO 1                                                 9
      P(I,1)=1.0/(1.0+EXP(-D*A(I,1)*(THETA-B(I,J-1))))                   10
      GO TO 2                                                            11
1     P(I,1)=0.0                                                         12
2     IF (J.EQ.(NCAT(I)+1)) GO TO 3                                      13
      P(I,2)=1.0/(1.0+EXP(-D*A(I,1)*(THETA-B(I,J))))                     14
      GO TO 4                                                            15
3     P(I,2)=1.0                                                         16
4     CONTINUE                                                           17
      RETURN                                                             18
      END                                                                19
      FUNCTION FDRVLL (IRESP,INAD,NITEMS,NCAT,A,B,THETA)                  1
      COMMON D                                                            2
      DIMENSION P(100,2), A(100,10), B(100,10), INAD(100), IRESP(100),'N  3
     1CAT(100)                                                            4
C***  CALCULATES FIRST DERIVATIVE OF LOG-LIKE FUNCTION                    5
      SUM=0.0                                                             6
      CALL CALCP (IRESP,INAD,NCAT,A,B,NITEMS,THETA,P)                     7
      DO 1 I=1,NITEMS                                                     8
      IF (INAD(I).EQ.0) GO TO 1                                           9
      SUM=SUM+A(I,1)*(1.0-P(I,1)-P(I,2))                                 10
1     CONTINUE                                                           11
      FDRVLL=SUM*D                                                       12
      RETURN                                                             13
      END                                                                14
      FUNCTION SDRVLL (IRESP,INAD,NITEMS,NCAT,A,B,THETA)                  1
      COMMON D                                                            2
      DIMENSION P(100,2), A(100,10), B(100,10), INAD(100), IRESP(100), N  3
     1CAT(100)                                                            4
C***  CALCULATES SECOND DERIVATIVE OF LOGLIKE FUNCTION                    5
      SUM=0.0                                                             6
      CALL CALCP (IRESP,INAD,NCAT,A,B,NITEMS,THETA,P)                     7
      DO 1 I=1,NITEMS                                                     8
      IF (INAD(I).EQ.0) GO TO 1                                           9
      Q1=1.0-P(I,1)                                                      10
      Q2=1.0-P(I,2)                                                      11
      P1=P(I,2)-P(I,1)                                                   12
      DIFF1=2.0*Q1-1.0                                                   13
      DIFF2=2.0*Q2-1.0                                                   14
      R1=P(I,1)*Q1                                                       15
      R2=P(I,2)*Q2                                                       16
      SUM=SUM+(A(I,1)**2/P1)*(-DIFF1*P1+DIFF2*P2-((R1-R2)**2)/P1)        17
1     CONTINUE                                                           18
      SDRVLL=D*D*SUM                                                     19
      RETURN                                                             20
      END                                                                21
      SUBROUTINE PCAT (A,B,THETA,ITEM,NUMCAT,PLOWER,PUPPER)               1
      COMMON D                                                            2
      DIMENSION B(100,10), PLOWER(10), PUPPER(10)                         3
C***  CALCULATES P'S FOR ALL RESPONSE CATEGORIES OF A GIVEN ITEM          4
C                                                                         5
      PLOWER(1)=0.0                                                       6
      DO 1 I=1,NUMCAT                                                     7
      PUPPER(I)=1.0/(1.0+EXP(-D*A*(THETA-B(ITEM,I))))                     8
      PLOWER(I+1)=PUPPER(I)                                               9
1     CONTINUE                                                           10
      PUPPER(NUMCAT+1)=1.0                                               11
      RETURN                                                             12
      END                                                                13
      SUBROUTINE LLINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)                1
      COMMON D                                                            2
      DIMENSION A(100,10), B(100,10), NCAT(100), INAD(100)                3
      DIMENSION PUPPER(10), PLOWER(10)                                    4
C***  COMPUTES INFORMATION FOR THE GRADED LOGISTIC MODEL AT A GIVEN       5
C***  VALUE OF THETA                                                      6
```

```
C                                                                       7
      TINFO=0.0                                                         8
      DO 2 I=1,NITEMS                                                   9
      IF (INAD(I).EQ.0) GO TO 2                                        10
      CALL PCAT (A(I,1),B,THETA,I,NCAT(I),PLOWER,PUPPER)               11
C                                                                      12
C***  LOOP OVER ALL THE RESPONSE CATEGORIES OF ITEM I                  13
      NCATEG=NCAT(I)+1                                                 14
      DO 1 J=1,NCATEG                                                  15
      QUPPER=1-PUPPER(J)                                               16
      QLOWER=1-PLOWER(J)                                               17
      P=PUPPER(J)-PLOWER(J)                                            18
      IF (P.EQ.0.0) P=0.0001                                           19
      FDRVP=D*A(I,1)*(PUPPER(J)*QUPPER-PLOWER(J)*QLOWER)               20
      TINFO=TINFO+(FDRVP*FDRVP)/P                                      21
1     CONTINUE                                                         22
2     CONTINUE                                                         23
      RETURN                                                           24
      END                                                              25
      SUBROUTINE NOMLOG (IRESP,A,B,NITEMS,NCAT,INAD,EPS,MAXIT,SFORM,IFAI  1
     1L,THETA,NUMITS,TINFO,SE)                                         2
      EXTERNAL FDRVNL                                                  3
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100) 4
      CALL BISECT (FDRVNL,IRESP,A,B,NITEMS,NCAT,INAD,5,GUESHI,GUESLO)  5
      CALL SECANT (FDRVNL,IRESP,A,B,NITEMS,NCAT,INAD,MAXIT,EPS,GUESHI,GU  6
     1ESLO,NUMITS,THETA,SDRV,IFAIL)                                    7
      SFORM=-SDRV                                                      8
      IF (IFAIL.EQ.1) GO TO 1                                          9
      CALL NLINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)                   10
      SE=1.0/SQRT(ABS(SDRV))                                           11
      RETURN                                                           12
C                                                                      13
1     TINFO=-99.99                                                     14
      SE=-99.99                                                        15
      RETURN                                                           16
      END                                                              17
      FUNCTION FDRVNL (IRESP,INAD,NITEMS,NCAT,A,B,THETA)                1
      DIMENSION IRESP(100), INAD(100), NCAT(100), A(100,10), B(100,10)  2
C***  CALCULATES FIRST DERIVATIVE OF NOMINAL LOGISTIC FUNCTION         3
C***  NOTE:  FOR THIS MODEL, THE A'S ARE THE SLOPE PARAMETERS          4
C***         AND THE B'S ARE THE INTERCEPT PARAMETERS                  5
C                                                                      6
      SUM=0.0                                                          7
      DO 2 I=1,NITEMS                                                  8
      IF (INAD(I).EQ.0) GO TO 2                                        9
      XNUM=0.0                                                        10
      DENOM=0.0                                                       11
      NUMCAT=NCAT(I)+1                                                12
      DO 1 J=1,NUMCAT                                                 13
      CONST=A(I,IRESP(I))-A(I,J)                                      14
      ARG=A(I,J)*THETA+B(I,J)                                         15
      EZ=EXP(ARG)                                                     16
      XNUM=XNUM+CONST*EZ                                              17
      DENOM=DENOM+EZ                                                  18
1     CONTINUE                                                        19
      SUM=SUM+XNUM/DENOM                                              20
2     CONTINUE                                                        21
      FDRVNL=SUM                                                      22
      RETURN                                                          23
      END                                                             24
      SUBROUTINE NLINFO (A,B,NITEMS,NCAT,INAD,THETA,TINFO)             1
      DIMENSION A(100,10), B(100,10), NCAT(100), INAD(100)             2
      DIMENSION E(10)                                                  3
C***  COMPUTES INFORMATION FOR THE NOMINAL LOGISTIC FUNCTION AT A      4
C***  GIVEN VALUE OF THETA                                             5
C                                                                      6
```

```
      TINFO=0.0                                                          7
C***  LOOP OVER THE ITEMS                                                8
      DO 4 I=1,NITEMS                                                    9
      IF (INAD(I).EQ.0) GO TO 4                                         10
      NCATEG=NCAT(I)+1                                                  11
C                                                                       12
C***  COMPUTE THE DENOMINATOR USED TO CALCULATE P AND THE FIRST         13
C***  DERIVATIVE OF P--IT IS THE SAME VALUE FOR ALL POSSIBLE RE-        14
C***  SPONSES TO ITEM I                                                 15
      DENOM=0.0                                                         16
      DO 1 K=1,NCATEG                                                   17
      E(K)=EXP(A(I,K)*THETA+B(I,K))                                     18
      DENOM=DENOM+E(K)                                                  19
1     CONTINUE                                                          20
C                                                                       21
C***  SUM OVER ALL POSSIBLE RESPONSES                                   22
      DO 3 J=1,NCATEG                                                   23
      P=E(J)/DENOM                                                      24
C                                                                       25
C***  CALCULATE THE FIRST DERIVATIVE OF P                               26
      DRVNUM=0.0                                                        27
      DO 2 K=1,NCATEG                                                   28
      DRVNUM=DRVNUM+E(K)*(A(I,J)-A(I,K))                                29
2     CONTINUE                                                          30
      FDRVP=(E(J)*DRVNUM)/(DENOM*DENOM)                                 31
C                                                                       32
      TINFO=TINFO+(FDRVP*FDRVP)/P                                       33
3     CONTINUE                                                          34
4     CONTINUE                                                          35
      RETURN                                                            36
      END                                                               37
      SUBROUTINE CHKRSP (IRESP,IOMIT,INAD,NCAT,NITEMS,NAME,ID,INADS)     1
      DIMENSION IRESP(100), INAD(100), NCAT(100), NAME(2), INADS(100)    2
C***  DELETES OMITTED ITEMS FOR A SUBJECT FROM LIST OF ITEMS ____        3
C***  ADMINISTERED (RESULTS RETURNED IN INADS)  AND CHECKS TO SEE        4
C***  IF ALL RESPONSES ARE VALID, IF NOT AN ERROR MESSAGE IS WRITTEN     5
C***  AND THE PROGRAM HALTS                                             6
C                                                                       7
      DO 2 I=1,NITEMS                                                   8
      NUMCAT=NCAT(I)+1                                                  9
      INADS(I)=INAD(I)                                                 10
      IF (IRESP(I).NE.IOMIT) GO TO 1                                   11
      INADS(I)=0                                                       12
      GO TO 2                                                          13
C                                                                      14
C***  CHECK FOR INVALID RESPONSE                                       15
1     IF ((1.LE.IRESP(I)).AND.(IRESP(I).LE.NUMCAT)) GO TO 2            16
      PRINT 3, NAME,ID,IRESP(I),NUMCAT                                 17
      STOP                                                             18
C                                                                      19
2     CONTINUE                                                         20
      RETURN                                                           21
C                                                                      22
C                                                                      23
3     FORMAT (10X,*ERROR IN DATA:  SUBJECT = *,2A10,* ID = *,A9,* HAS IL  24
     1LEGAL RESPONSE =*,I3,/10X,*RESPONSES MUST BE*,* BETWEEN 1 AND NUMB  25
     2ER OF CATEGORIES =*,I3,* INCLUSIVE*)                             26
      END                                                              27
      SUBROUTINE CHKVEC (IRESP,INAD,NCAT,NITEMS,ICHK,NBEST,NANS)        1
      DIMENSION IRESP(100), INAD(100), NCAT(100)                        2
C     CHKVEC TESTS THE RAW RESPONSE VECTOR TO DETERMINE IF              3
C     IT IS ONE OF THE FOLLOWING THREE CASES:                           4
C***     ALL BEST RESPONSES...ICHK SET TO 3                             5
C***     ALL WORST RESPONSES...ICHK SET TO 2                            6
C***     NO RESPONSES.....ICHK SET TO 1                                 7
C***     OTHERWISE.....ICHK SET TO 4                                    8
```

```
      NANS=0                                                              9
      NWORST=0                                                           10
      NBEST=0                                                            11
      DO 1 I=1,NITEMS                                                    12
      IF (INAD(I).EQ.0) GO TO 1                                          13
      NANS=NANS+1                                                        14
      IF (IRESP(I).EQ.1) NBEST=NBEST+1                                   15
      IF (IRESP(I).EQ.NCAT(I)+1) NWORST=NWORST+1                         16
1     CONTINUE                                                           17
C                                                                        18
      ICHK=4                                                             19
      IF (NANS.EQ.0) ICHK=1                                              20
      IF (NWORST.EQ.NANS) ICHK=2                                         21
      IF (NBEST.EQ.NANS) ICHK=3                                          22
      RETURN                                                             23
      END                                                                24
      SUBROUTINE BISECT (F1,IRESP,A,B,NITEMS,NCAT,INAD,NIT,BMID,BLAST)    1
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100),   2
     1P(100,2)                                                           3
C***  CALCULATES APPROXIMATE ROOT OF F1 BY BISECTION,BISECTING THE INTER  4
C***  NIT( NUMBER OF ITERATIONS) TIMES.  BMID IS BEST CURRENT GUESS AT T  5
C***  BLAST IS THE PREVIOUS VALUE OF BMID. IT IS USED AS THE SECOND       6
C***  INITIAL GUESS FOR THE SECANT METHOD.                               7
C                                                                        8
C     INITIALIZE LEFT BOUND AND F1(BOUND), AND RIGHT BOUND,F1(BOUND)     9
      BL=-5.0                                                           10
      BR=5.0                                                            11
      BMID=0.0                                                          12
      TL=F1(IRESP,INAD,NITEMS,NCAT,A,B,BL)                              13
      TR=F1(IRESP,INAD,NITEMS,NCAT,A,B,BR)                              14
C     TEST FOR NO ROOT IN INTERVAL,AND RETURN IF NOT POSSIBLE           15
      IF ((TL*TR).GT.0.0) RETURN                                        16
C     NOW CALCULATE BISECTIONS NIT TIMES                                17
      DO 3 I=1,NIT                                                      18
      TMID=F1(IRESP,INAD,NITEMS,NCAT,A,B,BMID)                          19
      IF ((TMID*TL).GT.0.0) GO TO 1                                     20
      BR=BMID                                                           21
      GO TO 2                                                           22
1     TL=TMID                                                           23
      BL=BMID                                                           24
2     BMID=(BL+BR)/2.0                                                  25
3     CONTINUE                                                          26
      BLAST=BR                                                          27
      IF (TL.EQ.TMID) BLAST=BL                                          28
      RETURN                                                            29
      END                                                               30
      SUBROUTINE NEWTRAP (F1,F2,IRESP,A,B,NITEMS,NCAT,INAD,NITER,EPS,NUM   1
     1ITS,GUESS,THETA,SDRV,IFAIL)                                         2
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100)    3
C     CALCULATES ROOT OF F1 GIVEN ITS FIRST DERIVATIVE F2 AND AN INITIAL  4
C***   GUESS.  UTILIZES NEWTON-RAPHSONS METHOD                           5
C                                                                        6
C     INITIALIZE                                                         7
      NUMITS=0                                                           8
      THETA=GUESS                                                        9
C***  LOOP UNTIL ERR IS LESS THAN EPS OR NUMBER OF ITERATIONS BECOMES TO  10
1     SDRV=F2(IRESP,INAD,NITEMS,NCAT,A,B,THETA)                          11
      FDRV=F1(IRESP,INAD,NITEMS,NCAT,A,B,THETA)                          12
      ERR=FDRV/SDRV                                                      13
      THETA=THETA-ERR                                                    14
      NUMITS=NUMITS+1                                                    15
C     EXIT LOOP CRITERION                                                16
      IF ((NUMITS.LT.NITER).AND.(ABS(ERR).GT.EPS)) GO TO 1               17
C***  END LOOP,  TEST FOR FAILURE AND SET IFAIL                          18
      IFAIL=0                                                            19
      IF (ABS(ERR).LT.EPS) RETURN                                        20
```

```
      IFAIL=1                                                          21
      THETA=-99.99                                                     22
      SDRV=99.99                                                       23
      RETURN                                                           24
      END                                                              25
      SUBROUTINE SECANT (F1,IRESP,A,B,NITEMS,NCAT,INAD,MAXIT,EPS,GUESHI,   1
     1GUESLO,NUMITS,THETA,SLOPE,IFAIL)                                 2
      DIMENSION IRESP(100), A(100,10), B(100,10), NCAT(100), INAD(100)  3
C***  USES THE SECANT METHOD TO CALCULATE THE ROOT, THETA, OF THE       4
C***  FUNCTION F1                                                      5
C***  GUESHI AND GUESLO ARE THE TWO INITIAL GUESSES AT THE ROOT        6
C***  REQUIRED BY THE SECANT METHOD                                    7
C                                                                      8
      NUMITS=0                                                         9
      THETA=GUESHI                                                     10
      TLAST=GUESLO                                                     11
      FLAST=F1(IRESP,INAD,NITEMS,NCAT,A,B,TLAST)                       12
C                                                                      13
C***  LOOP UNTIL CONVERGENCE OR NONCONVERGENCE IS ESTABLISHED           14
1     FCUR=F1(IRESP,INAD,NITEMS,NCAT,A,B,THETA)                        15
      IF (FCUR.EQ.FLAST) GO TO 2                                       16
      SLOPE=(THETA-TLAST)/(FCUR-FLAST)                                 17
      CHANGE=FCUR*SLOPE                                                18
      TLAST=THETA                                                      19
      FLAST=FCUR                                                       20
      THETA=THETA-CHANGE                                               21
      NUMITS=NUMITS+1                                                  22
      IF (ABS(CHANGE).GT.EPS.AND.NUMITS.LT.MAXIT) GO TO 1              23
C                                                                      24
      IFAIL=0                                                          25
      SLOPE=1.0/SLOPE                                                  26
      IF (ABS(CHANGE).LE.EPS) RETURN                                   27
C***  SECANT METHOD DOES NOT CONVERGE IN MAXIT ITERATIONS              28
      IFAIL=1                                                          29
      THETA=-99.99                                                     30
      SLOPE=-99.99                                                     31
      RETURN                                                           32
C                                                                      33
C***  ERROR:  SECANT METHOD CANNOT BE USED ON F1                       34
2     PRINT 3                                                          35
      IFAIL=1                                                          36
      THETA=-88.88                                                     37
      SLOPE=-88.88                                                     38
      RETURN                                                           39
C                                                                      40
3     FORMAT (1OX,*HORIZONTAL SLOPE FOUND IN FIRST DERIVATIVE FOR *,*CUR   41
     1RENT SUBJECT.*,/1OX,*SECANT METHOD CANNOT BE USED.*)             42
      END                                                              43
```

DISTRIBUTION LIST

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE
HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research
Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION &
TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Head, Research, Development, and Studies
(OP102X)
Office of the Chief of Naval Operations
Washington, DC 20370

1 Scientific Advisor to the Chief of
Naval Personnel (Pers-0r)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH CANTER
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

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

1 MCDEC
Quantico Marine Corps Base
Quantico, VA 22134

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

CoastGuard

1 MR. JOSEPH J. COWAN, CHIEF
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Vern W. Urry
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Werner Birke
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Progr
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Psychology
University of Illinois
Champaign, IL 61820

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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

1 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

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

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

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978

Research Reports

78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978. (NTIS No. AD A062797)
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978. (NTIS No. AD A059436)
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977. (NTIS No. AD A047495)
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977. (NTIS No. AD A046062)
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977. (NTIS No. AD A044828)
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977. (NTIS No. AD A041090)
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977. (NTIS No. AD A041084)
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977. (NTIS No. AD A038114)
Final Report: Computerized Ability Testing, 1972-1975. April 1976. (NTIS No. AD A024516)
76-5. Effects of Item Characteristics on Test Fairness. December 1976. (NTIS No. AD A035393)
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976. (NTIS No. AD A027170)
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976. (NTIS No. AD A028147)
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976. (NTIS No. AD A024422)
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (NTIS No. AD A022964)
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975. (NTIS No. AD A020961)
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (NTIS No. AD A018675)
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1976. (NTIS No. AD A018758)
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (NTIS No. AD A013185)
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975. (NTIS No. AD A007572)
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (NTIS No. AD A006733)
74-5. Strategies of Adaptive Ability Measurement. December 1974. (NTIS No. AD A004270)
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974. (NTIS No. AD A001230)
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (NTIS No. AD 783553)
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (NTIS No. AD 781894)
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974. (NTIS No. AD 773961)
73-3. The Stratified Adaptive Computerized Ability Test. September 1973. (NTIS No. AD 768376)
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973. (NTIS No. AD 757788).

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.

Copies of these reports are available, while supplies last, from:

Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

EFFECTS OF COMPUTERIZED ADAPTIVE TESTING ON BLACK AND WHITE STUDENTS

Steven M. Pine
Austin T. Church
Kathleen A. Gialluca
and
David J. Weiss

RESEARCH REPORT 79-2
MARCH 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

Prepared under contract No. N00014-76-C-0244, NR150-383
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Report 79-2	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Effects of Computerized Adaptive Testing on Black and White Students	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Steven M. Pine, Austin T. Church, Kathleen A. Gialluca, and David J. Weiss	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0244
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P.E.:61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.:NR150-383
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: March 1979
	13. NUMBER OF PAGES: 47
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 in 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)

ability tests	psychological reactions	motivation	ICC theory
adaptive tests	mode of administration	guessing	race
tailored tests	order of administration	bias	
conventional tests	knowledge of results	item bias	
bias-reduced tests	nervousness	test bias	

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

Bias-reduced and non-bias-reduced conventional paper-and-pencil and computerized adaptive tests of word knowledge were administered to Black and White high school students to study differential effects on ability estimates and psychological reactions. Independent variables examined were bias-reduction, the presence or absence of knowledge of results after each item, mode of administration (paper-and-pencil or computerized adaptive), order of administration, and race. Dependent variables were three test performance variables

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(the ability estimates derived from both conventional paper-and-pencil and computerized adaptive tests, the variance of those estimates, and the number of omitted responses) and four psychological reaction variables (reaction to knowledge of results, nervousness, motivation, and guessing). Bias-reduced tests were specially constructed from items which had previously been shown to be less biased towards Black students in terms of an item bias index derived from item characteristic curve (ICC) theory. The bias-reduced tests eliminated mean racial differences between Black and White students under certain test conditions, but the effect interacted with other conditions of test administration, e.g., whether or not knowledge of results was provided. Since the bias-reduced tests provided less precise measurement than the non-bias-reduced tests, it was concluded that more traditional item statistics, such as item discriminations, should be considered along with an index of item bias in test construction. Computerized adaptive tests were generally shown to be more motivating than the conventional paper-and-pencil tests. Black students, in particular, seemed to be less tolerant of the conventional paper-and-pencil tests, especially when taken after the adaptive test. This was reflected in levels of reported motivation, number of omitted responses, and reported amounts of guessing. Differential psychological reactions for Black and White students were found for other conditions of test administration as well; however, the computer-administered adaptive tests appeared to reduce these differences in comparison to the conventional paper-and-pencil tests. These data imply the need for further study of the effects of test administration conditions on members of minority groups to determine those administration conditions which maximize ability estimates directly or through their effects on the psychological environment of testing.

CONTENTS

Acknowledgements

We would like to thank the Minneapolis Public Schools for their approval and support of this research. We are particularly grateful to the students and staff of Central and North High Schools for their enthusiastic cooperation. Special thanks are also due the Association of Afro-American Educators for their support.

Technical Editor: Barbara Leslie Camm

Effects of Computerized Adaptive Testing on Black and White Students

Because computerized adaptive or tailored testing has the capability of individualizing ability tests to the characteristics of an examinee, it would appear to have the potential for reducing group differences in test scores resulting from individual or group difference variables other than those that the test is designed to measure. These variables might include group differences in motivation, test-taking anxiety, or tendency to guess or to omit items.

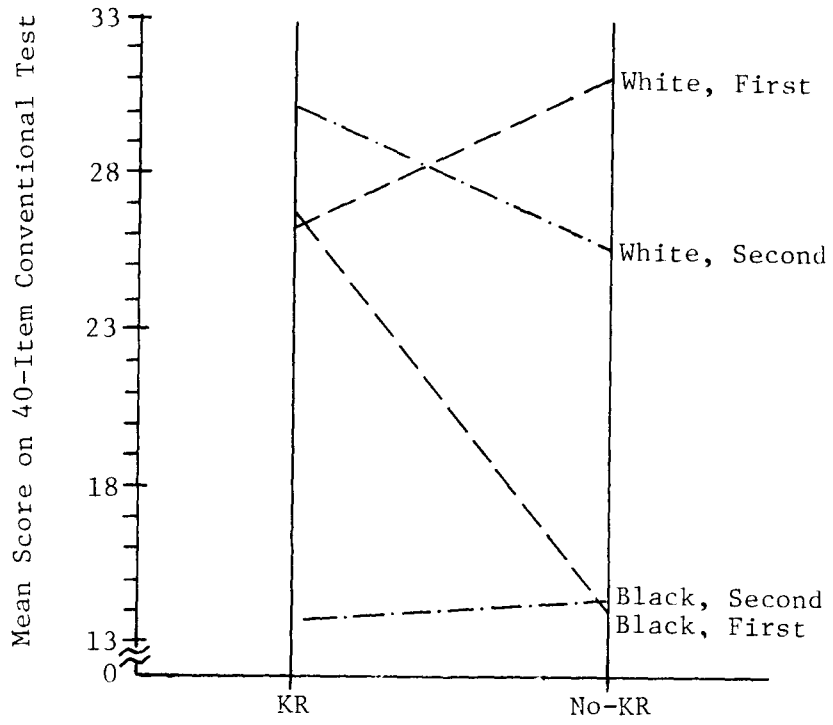
In conventional ability testing, items of the same difficulty are given to all examinees, regardless of their true ability levels. This reduces test reliability; consequently, the validity of the test may also be reduced in those groups which receive items inappropriate for their ability levels. Subgroups of the general population often differ with respect to background variables other than ability which may affect their performance on ability tests; therefore, test items which are appropriate in content for one subgroup may be inappropriate for another subgroup. With adaptive testing, it is possible to administer only those items that are appropriate for each group being tested. The process of adapting the test to each individual may also result in differential psychological impact on examinees from different population subgroups.

Previous research has provided some evidence for these potential psychometric and psychological benefits to minority examinees using computerized testing. Pine and Weiss (1978) demonstrated through a computer simulation that a Bayesian version of an adaptive test could reduce test unfairness within a simulated employee selection situation. In a live administration of computer-administered conventional tests, Johnson and Mihal (1973) administered identical conventional tests to Black and White students by paper and pencil and by computer. White students scored significantly higher than Black students on the paper-and-pencil tests, but not on the computer-administered tests.

In a study reported by Betz (1975, p. 24), two tests were administered by computer to a group of about 100 high school students, consisting of Black and White students. Both a conventional test and a pyramidal adaptive test (Larkin & Weiss, 1974) were administered to each student; half the group received the conventional test first, and half received the adaptive test first. In addition, half the group received feedback after each item indicating whether or not their answers were correct (knowledge of results, or KR, condition); the other half received no feedback after each test item (no knowledge of results, or No-KR, condition). The design was, therefore, a 2×2×2 analysis of variance. The independent variables were (1) race--Black and White, (2) knowledge of results (KR)--immediate or none, and (3) order--conventional test administered first or second. The data were analyzed for the conventional test only; thus, the dependent variable in this analysis was number-correct score on the conventional test.

The results for the three-way analysis of variance showed that the only significant main effect was for race. However, there was a significant three-way Order×Race×KR interaction. When a conventional test was administered first under conditions of immediate feedback, the mean of the Black students (26.4) was not significantly different from the mean of the White students (26.0), as is indicated in Figure 1.

Figure 1
Mean Scores for Black and White Students Completing a 40-Item Conventional Test First and Second in Both Knowledge of Results (KR) Conditions



If this result can be replicated, it implies that race differences observed in test scores may be a function, not of differences in ability levels, but of differences in the psychological effects of the conditions of administration. These findings, although not completely replicating those of Johnson and Mihal (1973), do support their general conclusion that conditions of test administration might affect motivational conditions, which in turn may reduce race group differences to nonsignificant levels.

The purpose of the present study was to replicate and to extend the previous findings that computerized administration of ability tests can increase the test scores and the test-taking motivation of minority examinees. Specifically, the present study compared a computerized adaptive test designed to minimize test bias with a similar conventional paper-and-pencil

test in order to investigate possible racial differences on the following variables:

1. Test performance variables
   a. Ability test scores
   b. Standard errors of measurement
   c. Number of omitted responses

2. Psychological reaction variables
   a. Reaction to knowledge of results
   b. Test-taking anxiety (nervousness)
   c. Motivation
   d. Tendency to guess.

METHOD

Subjects

Two hundred and thirty-four students from a Minneapolis high school were tested. Black and White students were about equally represented in the total group. A small amount of subject attrition occurred because of equipment failures and interruptions unrelated to the testing procedure, thus resulting in incomplete data sets. The number of missing subjects differed for different analyses and therefore is reported separately for each analysis. Each student was tested during the course of a normal school day and received a McDonald's gift certificate worth $.50 for participating in the study.

Design

The design for this study was a five-way factorial with repeated measures on one factor; the other four variables were completely crossed. Table 1 summarizes the five independent variables. Each student was assigned sequentially to one of the bias-reduction (BR)×knowledge of results (KR)×

Table 1
Description of Independent Variables

Independent Variable	Number of Conditions	Conditions	Type of Variable
Bias-Reduction (BR)	2	Bias-Reduced, Non-Bias-Reduced	Crossed
Knowledge of Results (KR)	2	Immediate Knowledge of Results, No Knowledge of Results	Crossed
Mode of Administration	2	Computer-Administered, Paper-and-Pencil	Repeated
Order of Administration	2	Paper-and-Pencil Test First, Computer-Administered Test First	Crossed
Race	2	Black, White	Crossed

order conditions within his/her respective racial group. The student was then administered two vocabulary tests--one conventional paper-and-pencil test and one computerized adaptive test--in the appropriate order. The major dependent variable derived from these tests was the student's ability level estimate obtained by scoring procedures based on item characteristic curve (ICC) theory. The number of omitted responses in each test was also recorded for each student. In addition to the vocabulary tests, each student was administered a test reaction questionnaire after each test condition.

Independent Variables

Bias Reduction

Item pool. The item pool consisted of 187 five-alternative multiple-choice word knowledge items gathered from several sources. Seventy-six of these items were developed and/or parameterized by Church, Pine, and Weiss (1978). Of these 76 items, 32 were written specifically as "Black-type" words; that is, it was assumed that the Black students would have greater familiarity with them than would the White students. Similarly, an additional 17 items were chosen as "White-type" words. Examples of each of these item types are given in Appendix Table A. The items not taken from Church et al. (1978) were obtained from the University of Minnesota computerized adaptive testing vocabulary item pool (McBride & Weiss, 1974).

For each item, item calibration procedures (see Church et al., 1978, pp. 19-22) yielded an index of bias and two standard ICC parameters (discriminating power, a, and item difficulty, b). The third ICC parameter, c, was set to .20 for all items, which is equal to 1 divided by the number of response alternatives. Bias was indexed by an ICC version of the Angoff and Ford (1971) elliptical distance measure of item bias (Martin, Pine, & Weiss, 1978). Since the elliptical distance index is highly correlated with the difference between the ICC item difficulties of the two contrasted groups, bias was indexed in the present study by the difference between the item difficulty (b) values for the Black and White groups. A positive value of the bias index indicates an item biased against the minority group, while a negative value indicates an item biased against the majority group. The calibrated item pool was then used to form two conventional paper-and-pencil tests and two computer-administered adaptive tests.

Computerized adaptive tests. The computer-administered adaptive tests (CAT) were constructed using the stradaptive testing strategy (Weiss, 1973). All items were assigned to one of seven strata based on the difficulty (b) parameter. Appendix Table B gives the a and b parameters and bias index for each item in the stradaptive pool.

To begin the stradaptive test, an initial stratum assignment was made by asking the students to rate themselves on verbal ability on a 3-point scale. Each student was asked the following question:

> Compared to other people, how good do you think your vocabulary is?
> 1. better than average, 2. average, 3. below average.

He/she was told to type a number from "1" to "3" accordingly. Students were then given the first item in Stratum 6, 4, or 2, depending on their

respective self-ratings. In accordance with usual stradaptive item selection procedures, students were subsequently administered items from the next-more-difficult or next-less-difficult stratum, depending on whether the response to the previous item was correct or incorrect. Each stradaptive test was terminated after 20 items.

Two forms of the adaptive test were constructed from the same item pool. In the bias-reduced (BR) adaptive test, items were arranged within each stratum in increasing order of bias. In the non-bias-reduced (NBR) adaptive test, items were arranged within each stratum in decreasing order of item discrimination, following recommendations for the construction of stradaptive tests (Weiss, 1974). Thus, in the BR condition, each item administered was the item with the lowest bias value still available in the appropriate stratum. In the NBR condition, each item administered was the most discriminating item remaining in the stratum.

Conventional paper-and-pencil tests. Two 20-item conventional paper-and-pencil (P&P) tests--one bias-reduced (BR) and one non-bias-reduced (NBR)--were constructed using items not used in the stradaptive test item pool. Item parameters and bias indices for these items are shown in Appendix Table C. The BR test included items with low positive or negative values of the bias index, while the NBR test included items with higher positive values of the bias index. Each set of 20 items formed a peaked test, with item difficulty peaked at the level of difficulty of Stratum 4, the middle stratum.

In order to equate conditions for the conventional paper-and-pencil and computer-administered adaptive tests as much as possible, items for the BR paper-and-pencil tests were selected to have approximately the same item bias values as the first few items that would be administered in each stratum of the BR adaptive test, and items for the NBR paper-and-pencil test were selected to have approximately the same item discrimination values as the first few items in each stratum of the NBR adaptive test. Consistent with this test-construction strategy, some items could be used in both the computerized tests and the paper-and-pencil tests as long as they were not in the same BR condition in both modes, since each student took the computerized and paper-and-pencil tests under only one BR condition.

It was impossible to match exactly the item characteristics of the 20 items in the conventional paper-and-pencil tests to the 20 items actually administered by the computerized adaptive tests, since it could not be determined in advance exactly which 20 items would be administered in the adaptive test to each student. Consequently, in order to compare these two testing strategies, the item characteristics of the computer-administered adaptive tests were calculated after administration of the tests (see Table 2 below).

Mode and Order of Administration

Each student completed a computer-administered test (adapted to his/her ability level) and a conventional paper-and-pencil test, both of which were either bias-reduced (BR) or non-bias-reduced (NBR). Half of the students took the paper-and-pencil test first (Order 1), while the other half took the computer-administered test first (Order 2).

The adaptive tests were computer administered by cathode-ray terminals (CRT) connected by telephone to a real-time computer system using procedures similar to those described by DeWitt and Weiss (1974). Each test item was presented separately on the CRT screen at the rate of 30 characters per second. Students were told that they could type a question mark in response to an item if they did not know the answer and wanted to omit it.

The paper-and-pencil tests were administered in booklets especially prepared for this study. Students had ample time to complete the tests and were instructed to omit an item if they did not know the correct answer.

Knowledge of Results

For half the students, immediate knowledge of results (KR) was administered after each test item, indicating whether or not the student's answer was correct; the other half received no information concerning the correctness of their answers (No-KR).

For the computer-administered tests, either the word *Correct* or *Incorrect* appeared on the screen after the student responded. The student then typed the letter *P* (for proceed) on the CRT keyboard in order to have the next question presented. In the No-KR condition, the next question appeared immediately after the student's answer was typed. KR in the paper-and-pencil mode was given using a latent ink process. Students marked their answer sheets with a special pen causing a latent image, which was previously invisible, to appear. The letter *Y* appeared if the correct answer was marked; the letter *N* appeared for incorrect answers.

Dependent Variables

Test Performance Measures

Three test performance measures were investigated. Ability level estimates were obtained using a Bayesian scoring procedure similar to the one developed by Owen (1975; see also McBride & Weiss, 1976, and Brown & Weiss, 1977, for applications of this ability estimation method). This scoring procedure provided a means of generating comparable scores for the conventional and adaptive tests. The posterior Bayesian variance, the second dependent variable used in this study, is the variance of the estimated ability score and can be considered an estimated standard error of estimate. The third dependent variable was the number of test questions omitted by each testee.

Psychological Reaction Scales

The psychological reactions to each condition were assessed by administering test reaction questions consisting of brief versions of four scales designed to assess reaction to knowledge of results, nervousness, motivation, and tendency to guess (see Betz & Weiss, 1976, for a description of the development of the scales from which these questions were selected). The test reaction questions are shown by scale in Appendix Table D along with the scaled scores used to obtain scores on the four scales. A student's score for each scale was the average of the scaled scores for the student's responses to the items in the scale.

The test reaction items were administered to each student twice, once after each test condition (computer-administered and paper-and-pencil). Students in the No-KR condition were given only the Nervousness, Motivation, and Guessing scales.

RESULTS

Test Characteristics

Test Items

To better interpret the meaning of any performance or motivational differences found between different testing conditions, it was important to examine the characteristics of the items administered under each testing condition. Because the computer-administered tests used a stradaptive strategy for item selection, it was not possible prior to administration to equate the item characteristics of the 20-item conventional paper-and-pencil tests to the 20-item computerized adaptive tests. As described earlier, items were divided between the paper-and-pencil and stradaptive item pools in order to equate, to the extent possible, item discriminations in the NBR condition and item bias in the BR condition.

Table 2 shows the mean, standard deviation, minimum and maximum values of item discrimination (a), difficulty (b), and bias parameters for the items in the conventional paper-and-pencil test and for the items actually administered in the computerized adaptive test under both BR and NBR conditions. For example, the average discrimination for items actually administered in the NBR adaptive test was 1.50, with discriminations of items administered ranging from about 1.00 to 2.27. These items also had a mean bias value of .72, indicating that the average item favored White students. For the conventional test in the NBR condition, the mean item discrimination was 1.57, with a range of 1.17 to 2.27.

Table 2
Item Discrimination (a), Difficulty (b), and Bias Values for the Conventional Paper-and-Pencil Tests and the Computerized Adaptive Tests in Bias-Reduced and Non-Bias-Reduced Conditions

Test and Statistic	Bias-Reduced (N=105)			Non-Bias-Reduced (N=106)		
	a	b	Bias	a	b	Bias
Conventional Test						
Mean	1.03	.02	-.05	1.57	.05	.83
S.D.	.47	.55	1.34	.28	.71	.36
Minimum	.09	-1.48	-5.46	1.17	-1.48	.22
Maximum	2.27	1.01	.74	2.27	1.46	1.71
Adaptive Test						
Mean	.84	-.10	-.20	1.50	-.41	.72
S.D.	.45	.91	1.22	.37	.59	.34
Minimum	.13	-1.61	-3.64	1.00	-1.51	.05
Maximum	1.96	2.04	1.29	2.27	.74	1.46

The data in Table 2 show that the strategy for item selection used in the BR condition did result in an adaptive test which was "bias-reduced," since the average bias value for items actually administered in the adaptive tests to students in the BR condition was -.20, which was lower than that for items administered in the NBR condition (mean = .72).

Not surprisingly, since NBR items were selected on the basis of their discrimination parameters, the average item administered in the adaptive test under the NBR condition was more discriminating (mean a=1.50) than the average item in the BR condition (mean a=.84). This was also reflected in the higher range of discrimination values in the NBR test.

In the conventional paper-and-pencil tests the item selection strategy resulted in the BR test having less "bias" against Black students (mean bias = -.05 compared to .83 in the NBR test), but it was also less discriminating (mean a=1.03 versus 1.57 for the NBR test). While the average item bias in the BR paper-and-pencil test favored Black students, examination of Appendix Table C indicates that this was attributable to a few items with large negative bias indices and that more of the items had small positive values of the bias index (i.e., favored White students). These items, however, had lower positive values of the bias index than most of the items in the NBR tests. Thus, while some of the items in the BR tests favored White students, the test items were, in general, more fair toward the Black students than the NBR tests.

Measurement Precision

Because increased item discrimination is related to increased item information, the NBR test might be expected to provide more precise ability estimates. In addition, previous research (Vale, 1975) has indicated that an adaptive test can yield more equiprecise measurement throughout the range of ability than a conventional peaked test. Using the Bayesian posterior variance as an estimate of the precision of measurement (Urry, 1977) at various levels of ability, Figures 2 and 3 provide support for both these expectations (numerical values for these figures are in Appendix Table E).

Figure 2 shows the mean Bayesian posterior variance for intervals of the Bayesian ability scores in the NBR condition; more precise measurement (lower posterior variance) was obtained with the adaptive test except for students whose ability level centered around the level of difficulty where the conventional test was peaked. In this range (θ=-.6 to .2) the conventional test had lower values of the Bayesian posterior variance.

Figure 3 shows the Bayesian posterior variance as a function of ability level for the BR condition for both adaptive and conventional tests. Under this test administration condition, items were selected by the adaptive test in order of their bias index, rather than by their discriminations. As Table 2 shows, the average discrimination of items administered in the adaptive test was lower than that in the conventional test. This is reflected in higher mean levels of the Bayesian posterior variance for the adaptive test for values of ability greater than θ=-1.00. In spite of this item selection procedure in the adaptive test, it still achieved lower average levels of the Bayesian posterior variance than did the conventional test for

Figure 2
Mean Bayesian Posterior Variance as a Function of Bayesian Ability Estimate for the Non-Bias-Reduced Adaptive and Conventional Tests

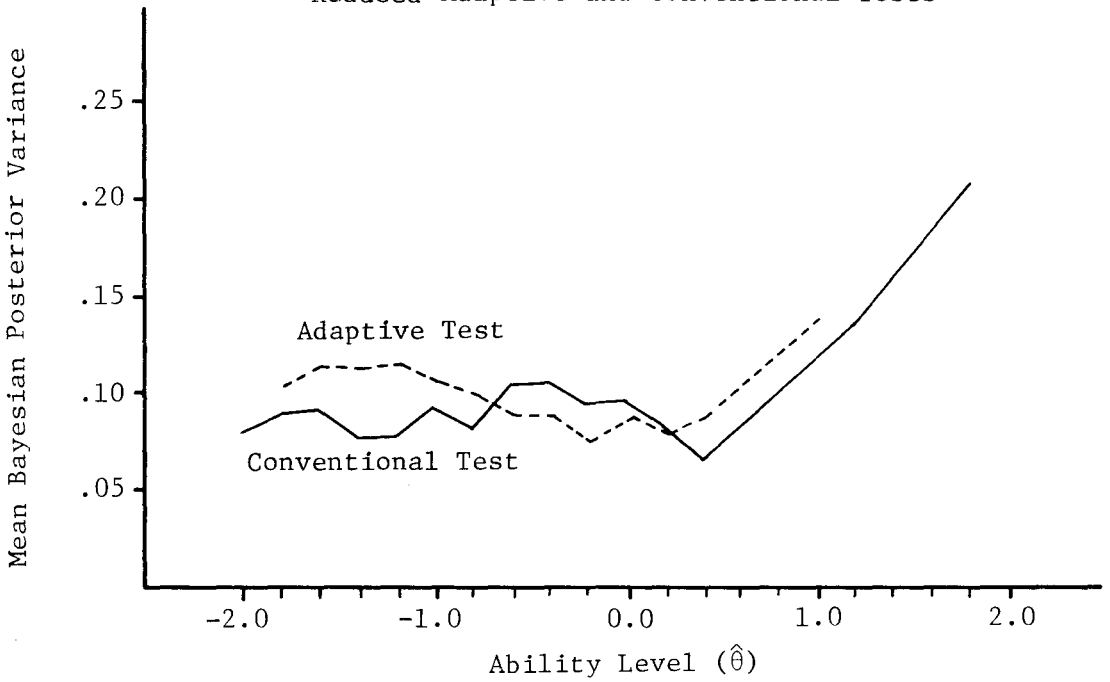
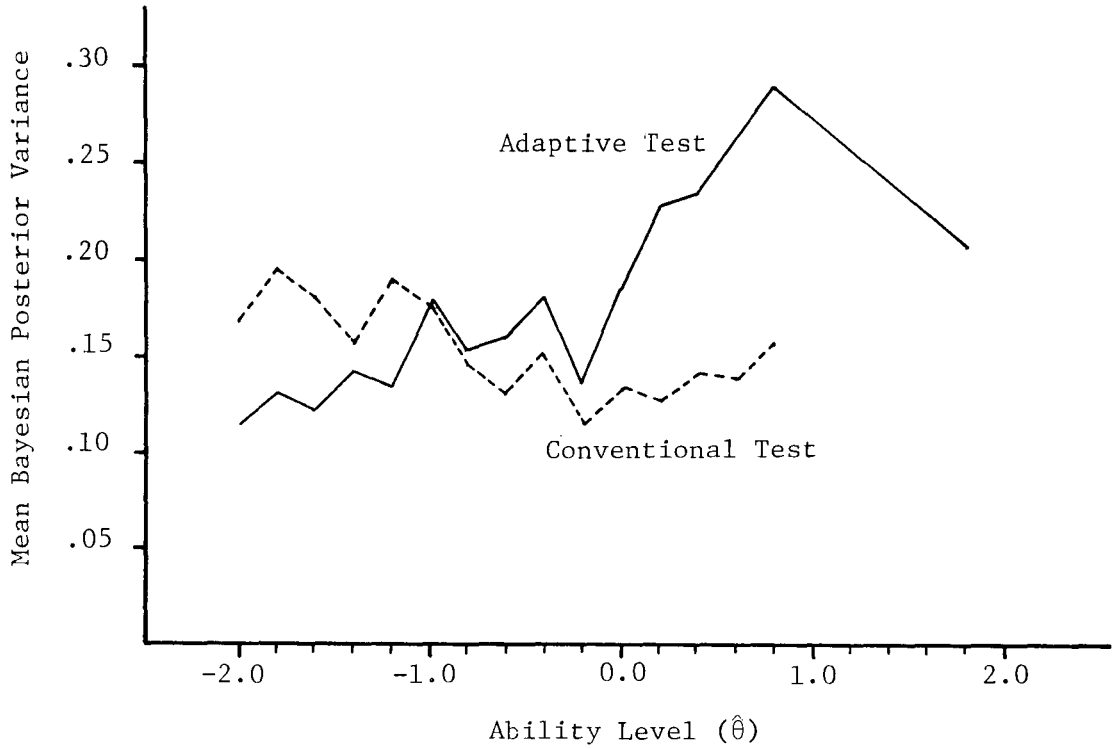


Figure 3
Mean Bayesian Posterior Variance as a Function of Bayesian Ability Estimate for the Bias-Reduced Conventional Test



ability levels less than θ=-1.00. The adaptive test compared more favorably with the conventional test in the NBR condition (Figure 2), however, supporting earlier recommendations that items within strata should be selected by their discrimination values when using a stradaptive testing strategy (Weiss, 1974).

Dependent Variables

Test Performance Variables

Appendix Tables F, G, and H show the means and standard deviations of the Bayesian ability estimates, Bayesian posterior variances, and number of omitted responses, respectively, for all combinations of the independent variables. Appendix Table I contains the means and standard deviations of these three dependent variables for various combined groups.

Ability estimates. The results of the 2×2×2×2×2 repeated measures analysis of variance for the Bayesian ability estimates are shown in Table 3. As this table indicates, the only statistically significant (p<.02) main effect was for race, with White students scoring higher (means = -.61 and -.63 for the computerized adaptive and conventional paper-and-pencil tests, respectively; see Table I) than Black students (means = -.87 and -.85, respectively). The interpretation of this significant main effect must be qualified, however, by a marginally significant three-way interaction between Race, KR, and BR (p=.07) and a four-way interaction between Mode, Race, KR, and BR (p<.06).

Figure 4 shows the four-way interaction (since it subsumes the three-way interaction) graphically by separately plotting the three-way interactions for both the computerized and paper-and-pencil administration modes. From this figure it can be seen that Black students did best in both testing modes when the test was bias-reduced and no knowledge of results was provided (BR, No-KR). In both tests this condition eliminated the main effect for race which existed in the other conditions. Black students obtained lowest mean scores (-1.02) in the paper-and-pencil test (Figure 4a) when the test was bias-reduced and knowledge of results was provided (BR, KR). On the computer-administered test in this condition (Figure 4b), mean score for the Black students was also relatively low.

The four-way interaction appeared to result primarily from the differential effect of the administration conditions on mean scores of the White students. As Figure 4a shows, highest mean scores were obtained for the White students on the paper-and-pencil test under the NBR and No-KR conditions. On the adaptive test (Figure 4b), however, the White students obtained lowest mean scores under these conditions. Comparison of Figures 4a and 4b also shows a general tendency for the adaptive test to reduce mean differences due to the interaction of race and testing conditions, since for both racial groups there was less variability among mean ability level scores as a function of testing conditions for the adaptive test, despite higher score variability (see Appendix Tables E and I).

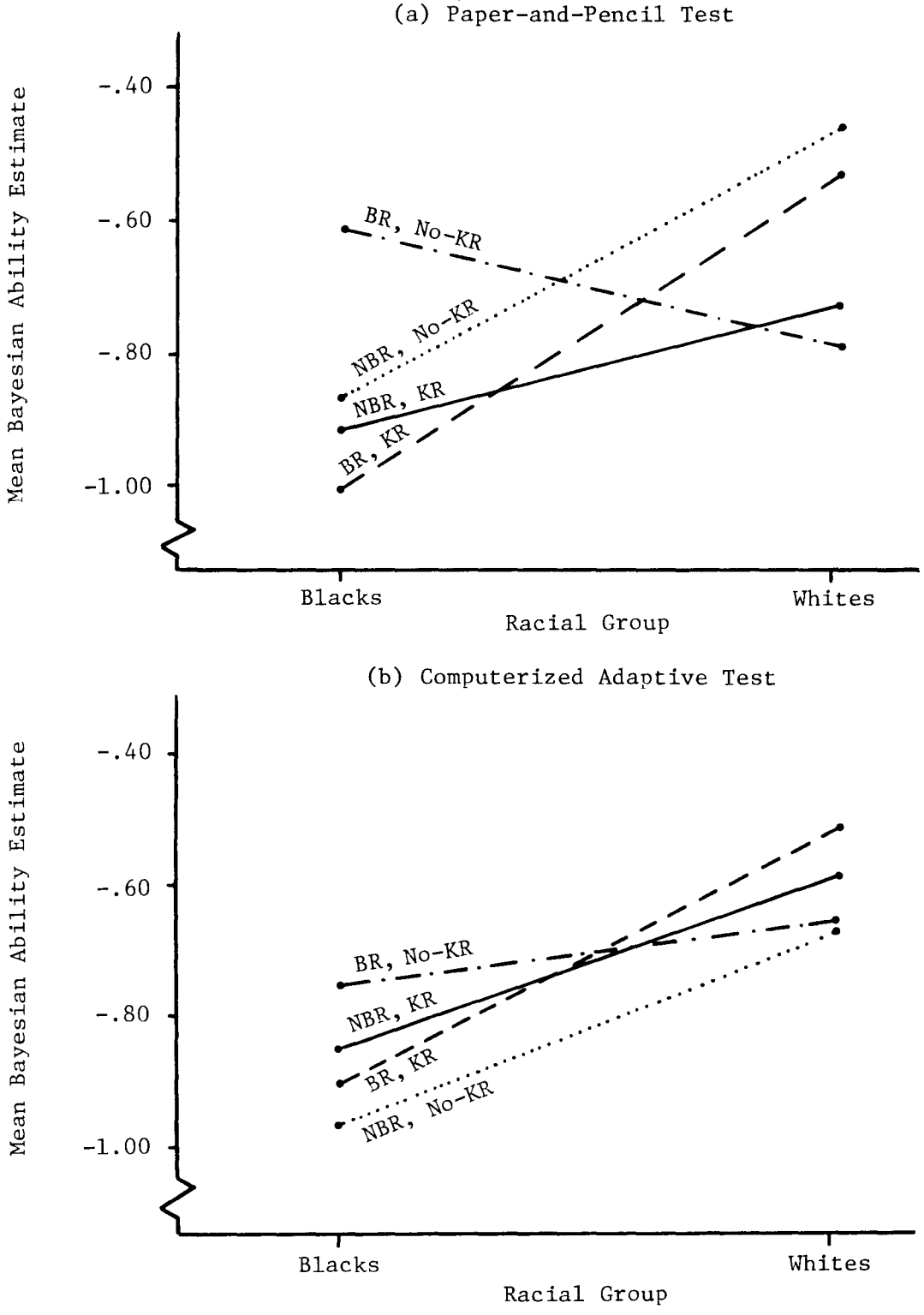
Consistency of ability estimates across modes. Of interest in comparing the computerized adaptive and conventional paper-and-pencil testing modes was

Table 3
Results of the Analysis of Variance for Bayesian Ability Estimates

Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	5.83	6.60	.001
Order	1	.95	1.08	.301
Knowledge of Results (KR)	1	.17	.19	.663
Bias Reduction (BR)	1	.09	.10	.754
Two-Way Interactions				
Race × Order	1	1.04	1.17	.280
Race × KR	1	.66	.75	.388
Order × KR	1	.00	.00	.982
Race × BR	1	.36	.41	.521
Order × BR	1	.19	.22	.640
KR × BR	1	.01	.02	.897
Three-Way Interactions				
Race × Order × KR	1	.01	.01	.910
Race × Order × BR	1	1.23	1.39	.240
Race × KR × BR	1	2.93	3.31	.070
Order × KR × BR	1	.49	.56	.456
Four-Way Interaction				
Race × Order × KR × BR	1	.51	.57	.450
Error	199	.88		
Within Subjects				
Main Effect				
Mode	1	.00	.02	.876
Two-Way Interactions				
Mode × Race	1	.14	.92	.338
Mode × Order	1	.10	.68	.409
Mode × KR	1	.43	2.83	.094
Mode × BR	1	.02	.17	.682
Three-Way Interactions				
Mode × Race × Order	1	.01	.08	.774
Mode × Race × KR	1	.21	1.40	.238
Mode × Order × KR	1	.01	.07	.799
Mode × Race × BR	1	.02	.15	.698
Mode × Order × BR	1	.43	2.84	.093
Mode × KR × BR	1	.14	.94	.333
Four-Way Interactions				
Mode × Race × Order × KR	1	.05	.30	.583
Mode × Race × Order × BR	1	.06	.39	.532
Mode × Race × KR × BR	1	.56	3.65	.057
Mode × Order × KR × BR	1	.44	2.86	.092
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.20	1.30	.255
Error	199	.15		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 4
Four-way Interaction of Mode of Administration, Race, Knowledge of Results (KR), and Bias Reduction (BR) for Bayesian Ability Estimates



the equivalence of the ability estimates obtained from the computerized and paper-and-pencil administrations. While the analyses of variance examined group level effects of test mode, it is also relevant to examine the similarity of rank orderings of individual student ability estimates across the two modes of test administration.

Pearson product-moment correlations between the ability estimates from the computer-administered adaptive test and the conventional paper-and-pencil test indicated substantial, but far from perfect, agreement between the two estimates for the sample as a whole (r=.73), for Black students (r=.70), for White students (r=.74), for students taking the BR tests (r=.72), and for students taking the NBR tests (r=.73). These correlations were all significantly different from zero (p<.01), but did not differ significantly from each other.

One probable reason for the moderate level of similarity of the ability estimates in the two modes of administration relates to the adaptive nature of the computer-administered tests. The distribution of students falling into various ability level intervals (see Appendix Table E), and the larger standard deviation of ability estimates in the adaptive test (*S.D.* = .80) as compared to the paper-and-pencil test (*S.D.* = .63), indicate that the adaptive test spread students out more on the ability continuum than did the conventional test. While ICC theory suggests that using Bayesian scoring ability estimates should not be dependent on the difficulty level of the items given, it appears that the peaked paper-and-pencil test was not able to locate people as well on the ability continuum if their ability levels were not near the point at which the test was peaked.

Bayesian posterior variance. Table 4 shows the results of the five-way repeated measures analysis of variance for the Bayesian posterior variance scores. A highly significant (p<.01) main effect for the bias-reduction factor was found, indicating that errors of measurement were larger in the BR tests (see Table I). This is consistent with the greater average discrimination of items in the NBR tests. The data in Table I also show that for the NBR tests, in which the adaptive test selected available items which were most discriminating, the adaptive test provided more precise ability estimates than the paper-and-pencil tests. For the BR tests, there was no advantage of the adaptive test over the paper-and-pencil tests in terms of accuracy of ability estimates. The bias-reduction factor was also involved, however, in the significant Race×Order×BR, Mode×BR, and Mode×Order×BR interactions. In addition, a significant Mode×Order effect was found.

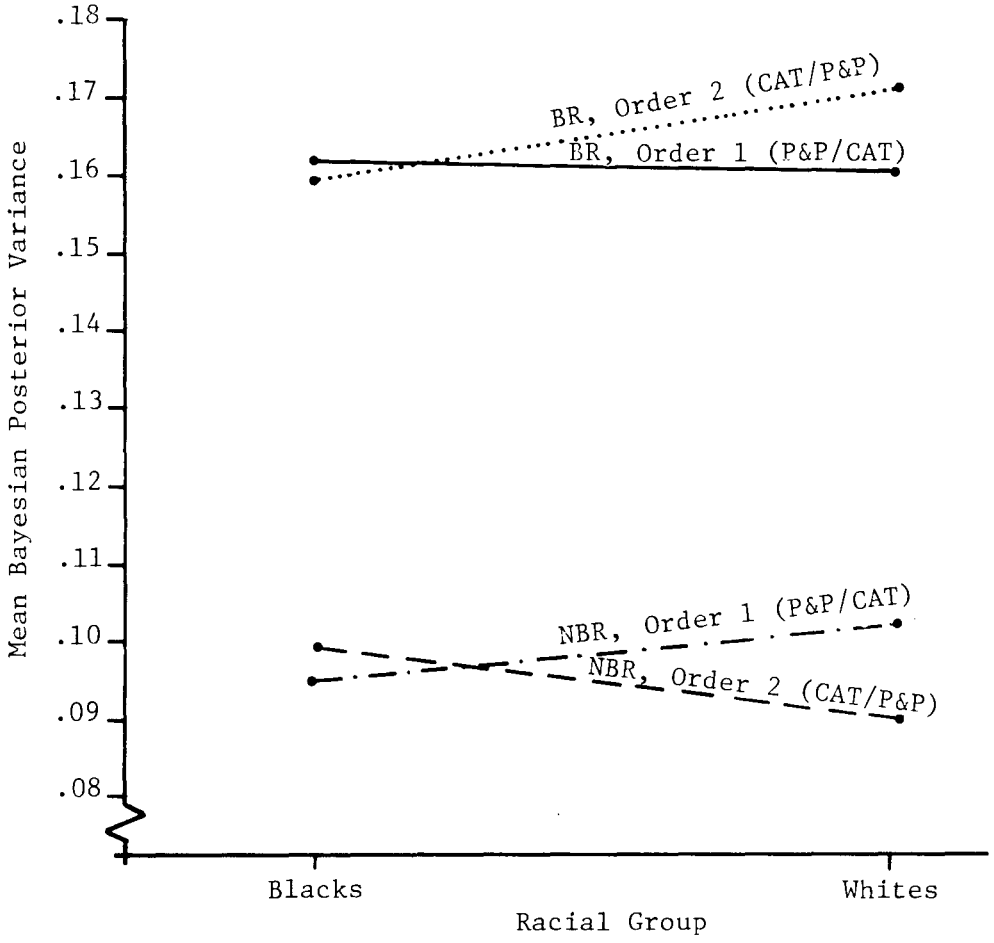
Figure 5 shows the Race×Order×BR three-way interaction. The increased precision obtained in the NBR conditions is clear in this figure, since lower values of the Bayesian posterior variance were obtained with the more highly discriminating items. The figure also shows that for the White group, posterior variances in the BR tests were smaller when the paper-and-pencil test was administered first (BR, P&P/CAT), while posterior variances were smaller in the NBR tests when the adaptive test was administered first (NBR, CAT/P&P). This pattern was reversed for Black students. In addition, the testing conditions had a greater effect on the Bayesian posterior variances for the White students.

Table 4
Results of the Analysis of Variance for Bayesian Posterior Variance Scores

Source of Variation	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	.00	.35	.554
Order	1	.00	.71	.401
Knowledge of Results (KR)	1	.00	3.06	.082
Bias Reduction (BR)	1	.45	582.28	.001
Two-Way Interactions				
Race × Order	1	.00	.066	.798
Race × KR	1	.00	1.58	.210
Order × KR	1	.00	.02	.893
Race × BR	1	.00	.67	.414
Order × BR	1	.00	2.78	.097
KR × BR	1	.00	2.30	.131
Three-Way Interactions				
Race × Order × KR	1	.00	3.47	.064
Race × Order × BR	1	.00	4.69	.031
Race × KR × BR	1	.00	.33	.564
Order × KR × BR	1	.00	.03	.870
Four-Way Interaction				
Race × Order × KR × BR	1	.00	.15	.697
Error	199	.00		
Within Subjects				
Main Effect				
Mode	1	.00	.50	.478
Two-Way Interactions				
Mode × Race	1	.00	2.16	.143
Mode × Order	1	.02	13.17	.001
Mode × KR	1	.00	.03	.872
Mode × BR	1	.01	6.01	.015
Three-Way Interactions				
Mode × Race × Order	1	.00	1.50	.222
Mode × Race × KR	1	.00	.16	.691
Mode × Order × KR	1	.00	.62	.430
Mode × Race × BR	1	.00	.67	.413
Mode × Order × BR	1	.01	9.32	.003
Mode × KR × BR	1	.00	1.74	.188
Four-Way Interactions				
Mode × Race × Order × KR	1	.00	.54	.463
Mode × Race × Order × BR	1	.00	2.12	.147
Mode × Race × KR × BR	1	.00	.00	.959
Mode × Order × KR × BR	1	.00	.12	.725
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.00	.04	.849
Error	199	.00		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 5
Three-Way Interaction of Race, Order of Administration, and Bias-Reduction (BR) for Bayesian Posterior Variance

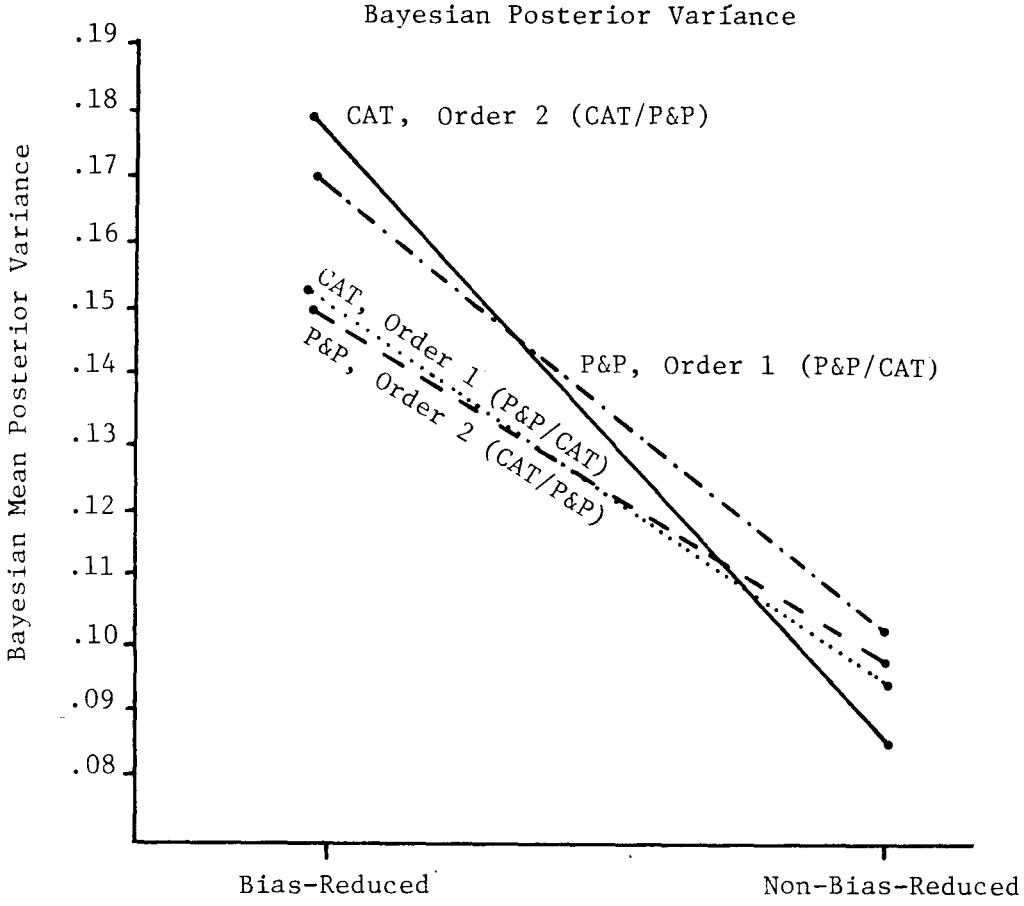


All the other significant interactions for the Bayesian posterior variance measure were subsumed in the Mode×Order×BR three-way interaction shown in Figure 6.

These data show that the combination of bias-reduced administration and order of administration affected Bayesian posterior variances on the adaptive test. Specifically, when the adaptive test was administered first (Order 2), it had the highest average level of the posterior variance among all test administration conditions in the BR condition and the lowest level in the NBR condition.

Number of omitted responses. Table 5 shows the results of the analysis of variance for the number of omitted responses. These data indicate a statistically significant ($p<.02$) main effect for KR, with students omitting more responses when KR was not given (see Table I). Examination of the statistically significant ($p<.05$) two-way interaction of the KR variable with the race factor (see Figure 7), however, indicates that this effect of KR on the number of omitted responses was largely due to its effect on the Black students.

Figure 6
Three-Way Interaction of Mode of Administration, Order of Administration, and Bias-Reduction (BR) for Bayesian Posterior Variance



As Figure 7 indicates, KR had a differential effect on the Black students, but no effect on the White students. When KR was administered, Black students omitted fewer items (mean = 1.88) than when KR was not given (mean = 3.89). In comparison, White students omitted an average of 2.75 and 2.68 items under KR and No-KR conditions, respectively.

The only other statistically significant interaction for omitted responses was the three-way interaction of Mode×Race×Order ($p<.05$). This interaction, pictured in Figure 8, shows that Black and White students differed in the relative number of responses they omitted on the paper-and-pencil test depending on whether that test was taken first or second. For the Black students, the highest mean number of omitted responses as a group occurred when the paper-and-pencil test was taken second (Order 2); and the fewest, when this test was taken first (Order 1). For the White students, the mean number of omitted responses on the paper-and-pencil test was highest when this test was given first (Order 1) and fewest when this test was given second (Order 2). In addition, the test administration variables resulted in slightly greater mean differences for the White students than for the Black students.

Table 5
Results of the Analysis of Variance for Number of Omitted Responses

Source of Variation	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	.25	.01	.914
Order	1	14.08	.65	.419
Knowledge of Results (KR)	1	123.64	5.76	.017
Bias Reduction (BR)	1	7.20	.33	.563
Two-Way Interactions				
Race × Order	1	14.93	.69	.405
Race × KR	1	99.18	4.62	.033
Order × KR	1	69.35	3.23	.074
Race × BR	1	.00	.00	.993
Order × BR	1	8.00	.37	.542
KR × BR	1	39.93	1.86	.174
Three-Way Interactions				
Race × Order × KR	1	.75	.03	.852
Race × Order × BR	1	3.90	.18	.671
Order × KR × BR	1	37.89	1.76	.186
Four-Way Interaction				
Race × Order × KR × BR	1	1.72	.08	.777
Error	206	21.48		
Within Subjects				
Main Effect				
Mode	1	.48	.06	.811
Two-Way Interactions				
Mode × Race	1	.47	.06	.813
Mode × Order	1	.02	.00	.956
Mode × KR	1	10.33	1.24	.267
Mode × BR	1	.26	.03	.859
Three-Way Interactions				
Mode × Race × Order	1	38.53	4.63	.033
Mode × Race × KR	1	23.71	2.85	.093
Mode × Order × KR	1	5.71	.68	.409
Mode × Race × BR	1	5.34	.64	.424
Mode × Order × BR	1	18.04	2.17	.143
Mode × KR × BR	1	2.34	.28	.596
Four-Way Interactions				
Mode × Race × Order × KR	1	.45	.05	.816
Mode × Race × Order × BR	1	8.70	1.04	.308
Mode × Race × KR × BR	1	22.37	2.69	.103
Mode × Order × KR × BR	1	3.68	.44	.507
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.38	.05	.830
Error	206	8.32		

*Estimated probability of error in rejection of the null hypothesis of no mean difference.

Figure 7
Two-Way Interaction of Race and Knowledge of Results (KR) for Number of Omitted Responses

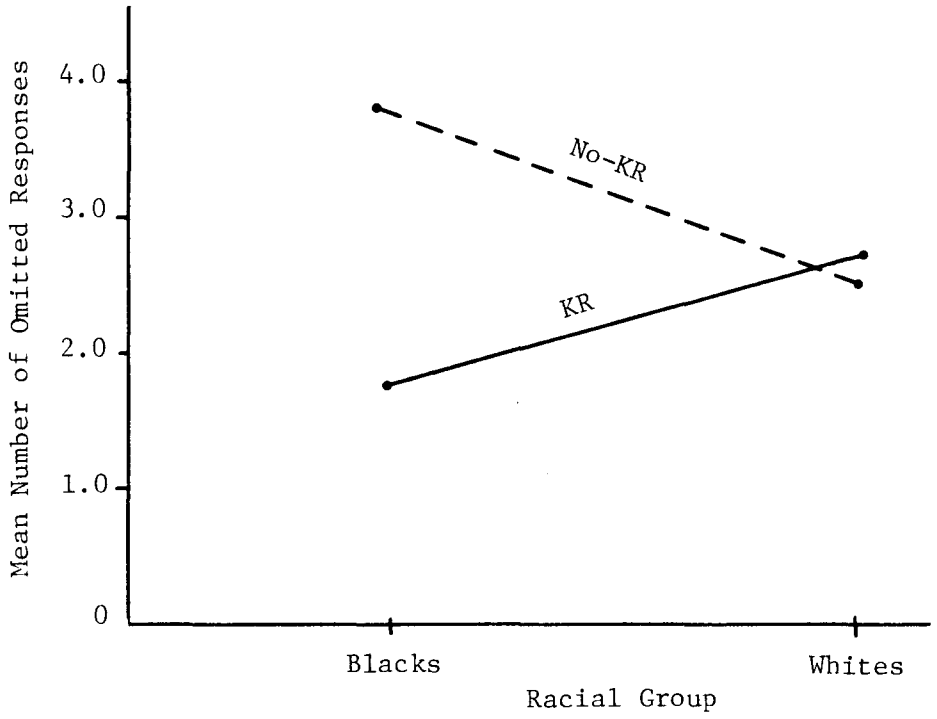
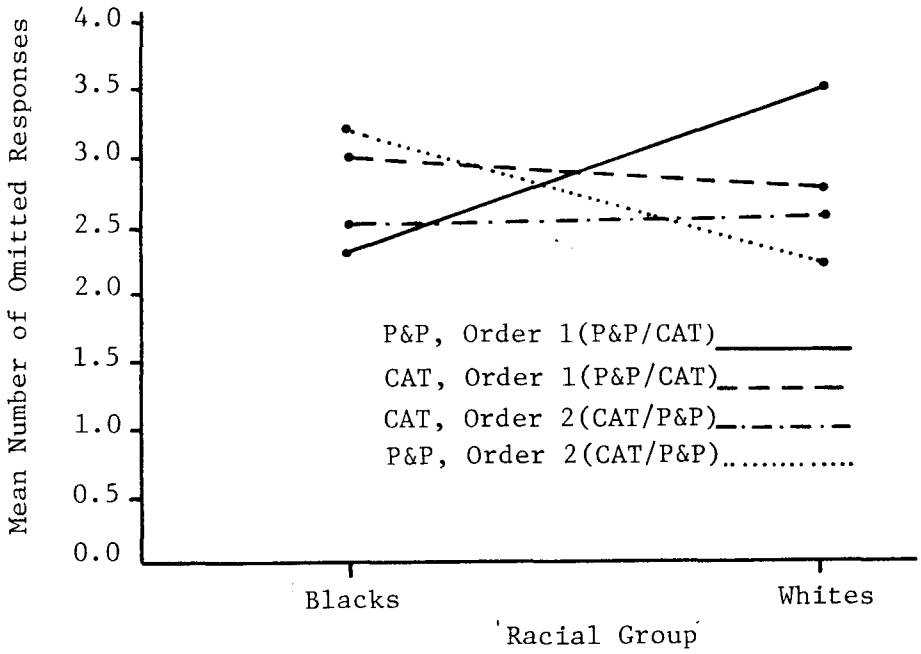


Figure 8
Three-Way Interaction of Mode of Administration, Race, and Order of Administration for Number of Omitted Responses



Psychological Reaction Variables

Means and standard deviations of the psychological reactions scales for all experimental conditions are in Appendix Tables J, K, L, and M, respectively, for the Knowledge of Results, Nervousness, Motivation, and Guessing scales. The means and standard deviations of the four psychological test reactions scales for the combined Racial, Bias-Reduction, Knowledge of Results, Order of Administration, and Mode of Administration groups are given in Appendix Table N.

Knowledge of Results. Table 6 gives the results of the analysis of variance of the scores on the reaction to Knowledge of Results scale. There was a statistically significant (p=.001) effect for race in the ANOVA of the reaction to Knowledge of Results scores, with Black students scoring higher on this scale than White students. This indicated a more negative attitude toward receiving KR after each item on the part of the Black students, i.e., they were more inclined to report that receiving KR made them nervous and interfered with their concentration.

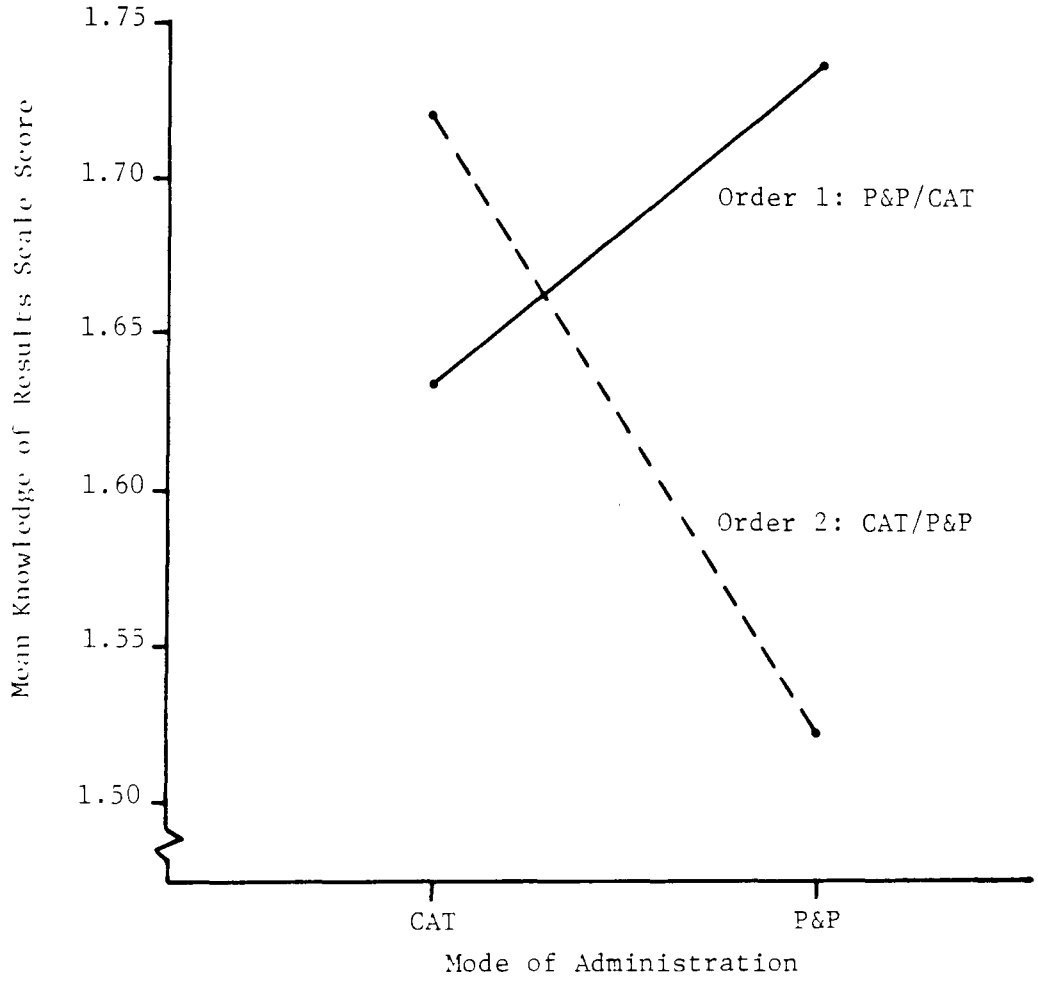
Table 6
Results of the Analysis of Variance of the Knowledge of Results Scale Scores

Source of Variance	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	11.25	12.59	.001
Order	1	.39	.44	.509
Bias Reduction (BR)	1	.06	.06	.804
Two-Way Interactions				
Race × Order	1	.96	1.08	.302
Race × BR	1	.10	.11	.742
Order × BR	1	1.24	1.39	.242
Three-Way Interaction				
Race × Order × BR	1	.92	1.03	.313
Error	88	.89		
Within Subjects				
Main Effect				
Mode	1	.29	1.42	.236
Two-Way Interactions				
Mode × Race	1	.00	.02	.899
Mode × Order	1	.94	4.63	.034
Mode × BR	1	.07	.35	.558
Three-Way Interactions				
Mode × Race × Order	1	.27	1.31	.256
Mode × Race × BR	1	.48	2.34	.129
Mode × Order × BR	1	.23	1.13	.290
Four-Way Interaction				
Mode × Race × Order × BR	1	.19	.91	.342
Error	88	.20		

*Estimated probability of error in rejecting the null hypothesis of no difference in group means.

The Mode×Order interaction was also statistically significant ($p<.05$) and is illustrated in Figure 9. Students reported a more favorable attitude toward KR (i.e., lower mean scale scores) during the second test than during the first. This was particularly true when the paper-and-pencil test was administered second, which was the condition resulting in the most favorable reactions to KR. The data in Figure 9 also show that students' reactions to computer-administered KR were less affected by the order of its administration than was paper-and-pencil-administered KR.

Figure 9
Two-Way Interaction of Mode of Administration and Order of Administration for the Knowledge of Results Scale Scores



Nervousness. The means and standard deviations of responses on the Nervousness scale are reported in Appendix Tables K and N; Table 7 gives the results of the analysis of variance for this scale.

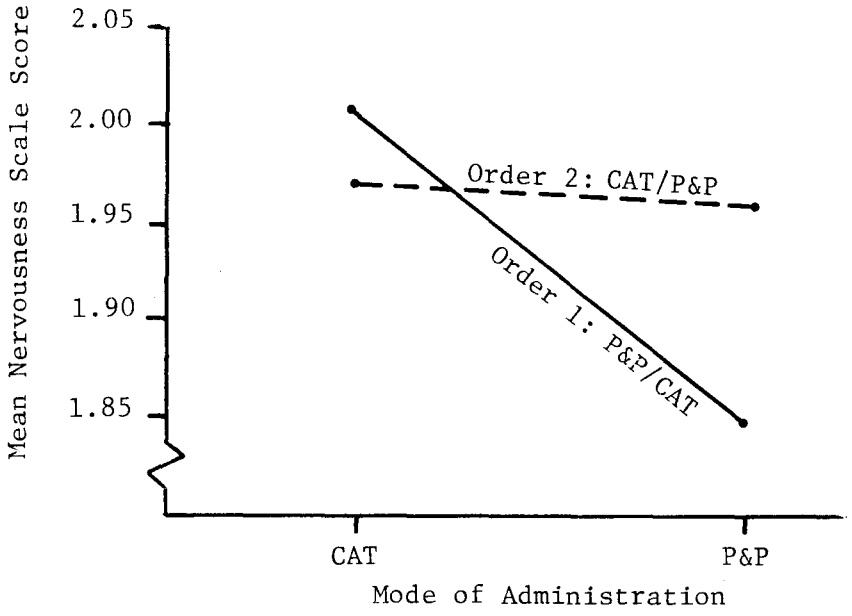
The only main effect that emerged as statistically significant ($p<.05$) was that of mode of administration, in which students reported that they were more nervous while taking the computer-administered test (mean = 2.02; see Table N) than they were while taking the paper-and-pencil test (mean = 1.91). The Mode×Order interaction was marginally significant ($p=.076$) and is shown in Figure 10. This figure shows that students reported lowest levels of nervousness

Table 7
Results of the Analysis of Variance of Nervousness Scale Scores

Source of Variation	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	.65	.90	.344
Order	1	.01	.01	.909
Knowledge of Results (KR)	1	1.58	2.19	.140
Bias Reduction (BR)	1	.60	.83	.364
Two-Way Interactions				
Race × Order	1	1.34	1.86	.174
Race × KR	1	.18	.25	.619
Race × BR	1	1.44	2.00	.159
Order × KR	1	.14	.19	.663
Order × BR	1	5.37	7.45	.007
KR × BR	1	.57	.79	.376
Three-Way Interactions				
Race × Order × KR	1	1.01	1.40	.238
Race × Order × BR	1	.38	.53	.468
Race × KR × BR	1	.27	.38	.540
Order × KR × BR	1	.09	.13	.717
Four-Way Interaction				
Race × Order × KR × BR	1	3.17	4.41	.037
Error	185	.72		
Within Subjects				
Main Effect				
Mode	1	1.22	5.01	.026
Two-Way Interactions				
Mode × Race	1	.12	.50	.480
Mode × Order	1	.78	3.19	.076
Mode × KR	1	.07	.30	.584
Mode × BR	1	.87	3.57	.060
Three-Way Interactions				
Mode × Race × Order	1	.01	.03	.868
Mode × Race × KR	1	.01	.05	.825
Mode × Race × BR	1	.16	.66	.416
Mode × Order × KR	1	.14	.56	.455
Mode × Order × BR	1	.16	.64	.423
Mode × KR × BR	1	.01	.02	.825
Four-Way Interactions				
Mode × Race × Order × KR	1	.00	.00	.985
Mode × Race × Order × BR	1	.15	.61	.435
Mode × Race × KR × BR	1	.02	.09	.760
Mode × Order × KR × BR	1	.04	.16	.689
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.30	1.25	.265
Error	185	.24		

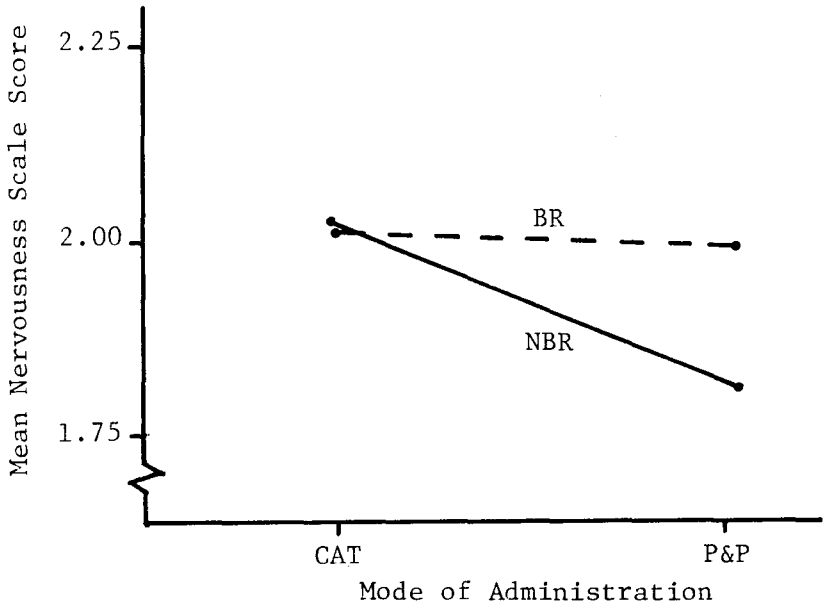
*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 10
Two-Way Interaction of Mode of Administration and
Order of Administration for Nervousness Scale Scores



when the paper-and-pencil test (Order 1) was administered first in a pair of tests (mean = 1.85) and highest levels when they were subsequently transferred to the computerized adaptive test (mean = 2.06). However, when students were first administered the computerized test (Order 2), their reported levels of nervousness remained about the same across both tests.

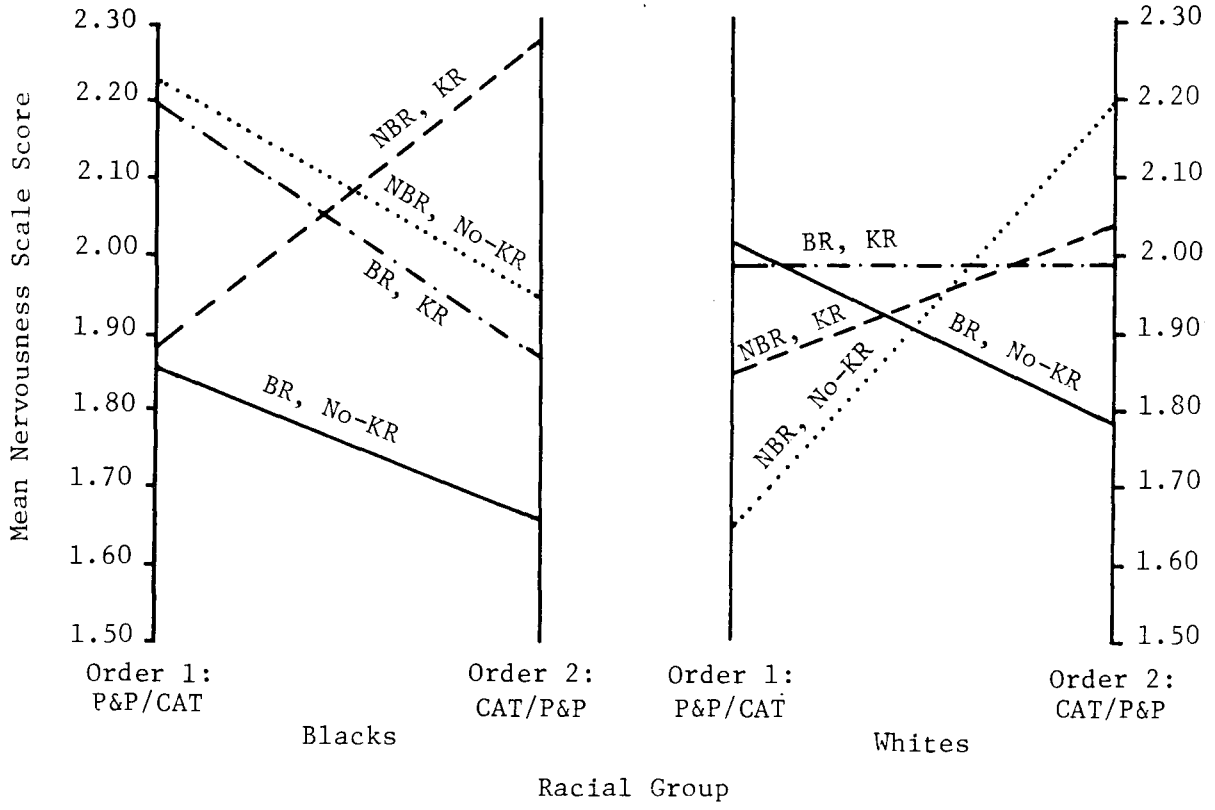
Figure 11
Two-Way Interaction of Mode of Administration
and Bias-Reduction (BR) for Nervousness Scale Scores



The Mode×BR interaction was also marginally significant (p=.06). Inspection of the graph of this interaction (Figure 11) indicates that the students reported equal levels of nervousness in both BR and NBR tests when they were administered adaptively by computer. When tests were administered by paper-and-pencil, however, lower levels of nervousness were observed in the BR condition.

There was also a statistically significant (p=.007) Order×BR interaction. Interpretation of this interaction is complicated by the presence of a four-way Race×Order×BR×KR interaction (p=.037), which is shown in Figure 12. As Figure 12 shows, reported nervousness of Black and White students was differentially affected by the Order, KR, and BR test administration conditions. Black students reported lower levels of nervousness when the computerized adaptive test was administered first if the tests were administered in the BR mode (with or without KR) and when the NBR test was administered without KR; they reported highest levels of nervousness when the NBR adaptive test was administered first with KR. For the Black students, lowest levels of nervousness were reported in the BR, No-KR condition, regardless of test order. For the White students, order of administration did not affect their reported nervousness in the BR, KR condition; the NBR, No-KR condition resulted in lowest levels of reported nervousness when the paper-and-pencil test was

Figure 12
Four-Way Interaction of Race, Order of Administration, Knowledge of Results (KR), and Bias-Reduction (BR) for Nervousness Scale Scores



administered first and highest levels of nervousness when it was administered second. Order of administration also affected the White students in opposite ways under the other two test administration condition combinations.

Motivation. The means and standard deviations of responses on the Motivation scale are given in Appendix Tables L and N; results of the analysis of variance for this scale are given in Table 8. Again, there was a statistically significant ($p<.01$) main effect for mode of administration, with students reporting that they were more motivated to perform well when they were taking the computer-administered test (mean = 2.99; see Table N) than when they took the paper-and-pencil test (mean = 2.86).

The Mode×Order interaction was marginally significant ($p=.071$) for this scale, but it was subsumed in the significant ($p=.022$) four-way Mode×Order×Race×BR interaction. The two-way Mode×BR and Race×Order interactions were also statistically significant ($p=.005$ and .021, respectively); these were also subsumed in the significant Race×Order×Mode×BR interaction, which is shown in Figure 13.

Figure 13
Four-Way Interaction of Mode of Administration, Race, Order of Administration, and Bias-Reduction (BR) for Motivation Scale Scores

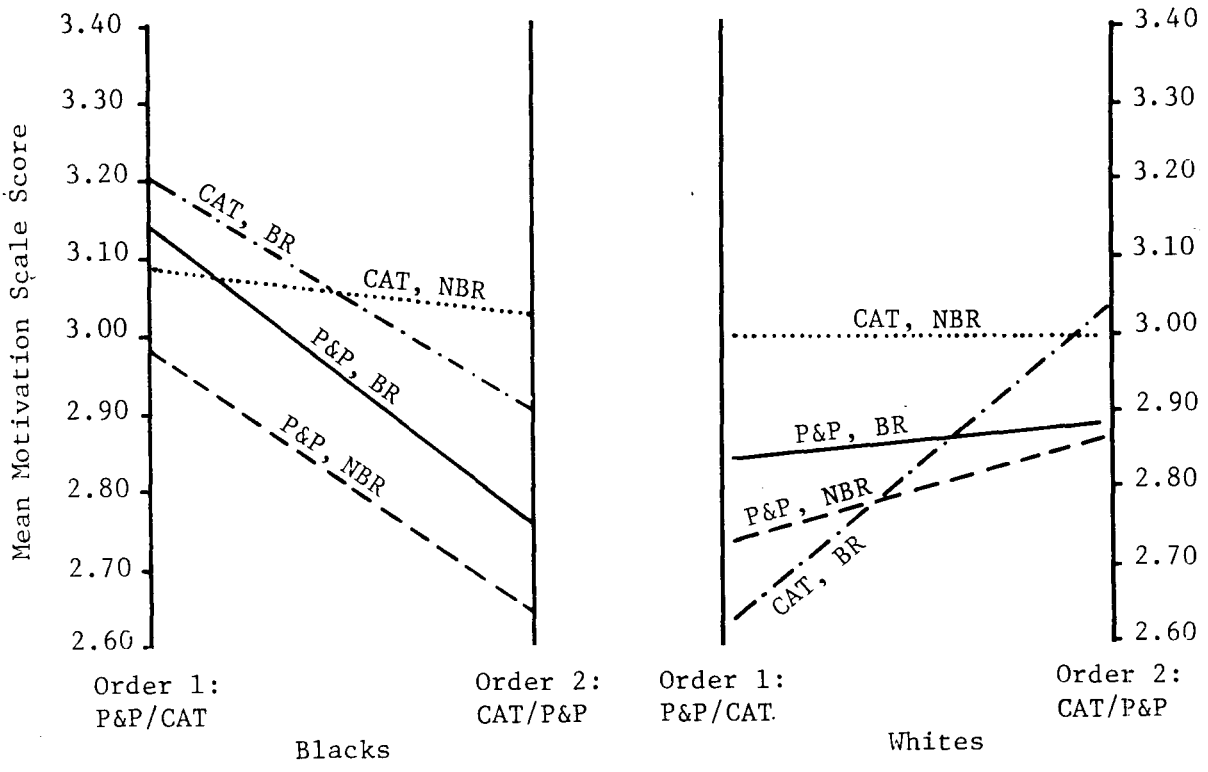


Figure 13 shows that reported motivation was uniformly lower for Black students in Order 2 (CAT/P&P) than in Order 1 (P&P/CAT). However, Order 2 had a greater effect on motivation reported after the paper-and-pencil test administration than after administration of the adaptive test. For the Black

Table 8
Results of the Analysis of Variance for Motivation Scale Scores

Source of Variation	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	1.00	1.16	.283
Order	1	.22	.25	.616
Knowledge of Results (KR)	1	.20	.23	.628
Bias Reduction (BR)	1	.00	.00	.997
Two-Way Interactions				
Race × Order	1	4.68	5.41	.021
Race × KR	1	.21	.24	.624
Race × BR	1	.29	.34	.562
Order × KR	1	.79	.91	.340
Order × BR	1	.02	.03	.867
KR × BR	1	.16	.19	.665
Three-Way Interactions				
Race × Order × KR	1	1.69	1.96	.164
Race × Order × BR	1	.39	.44	.506
Race × KR × BR	1	.92	1.07	.303
Order × KR × BR	1	2.34	2.70	.102
Four-Way Interaction				
Race × Order × KR × BR	1	5.10	5.89	.016
Error	185	.87		
Within Subjects				
Main Effect				
Mode	1	2.17	14.04	.000
Two-Way Interactions				
Mode × Race	1	.09	.59	.445
Mode × Order	1	.51	3.31	.071
Mode × KR	1	.25	1.64	.202
Mode × BR	1	1.25	8.09	.005
Three-Way Interactions				
Mode × Race × Order	1	.02	.16	.689
Mode × Race × KR	1	.31	2.01	.158
Mode × Race × BR	1	.21	1.33	.251
Mode × Order × KR	1	.25	1.60	.208
Mode × Order × BR	1	.13	.84	.360
Mode × KR × BR	1	.01	.08	.783
Four-Way Interactions				
Mode × Race × Order × KR	1	.17	1.10	.297
Mode × Race × Order × BR	1	.83	5.35	.022
Mode × Race × KR × BR	1	.01	.04	.833
Mode × Order × KR × BR	1	.38	2.47	.118
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.03	.20	.654
Error	185	.15		

*Estimated probability of error in rejecting the null hypothesis of no mean differences.

students, lowest levels of motivation in both orders of administration were reported for the NBR paper-and-pencil test; highest levels of reported motivation were reported in Order 1 on the BR adaptive test. In general, order of administration had an opposite effect on White students; reported levels of motivation were higher for Order 2 than for Order 1. For Whites, the BR adaptive test resulted in lowest levels of reported motivation when it was administered second and highest levels when it was administered first. For both the Black and White groups, the NBR adaptive test was the only testing condition for which order of administration did not affect reported motivation.

Guessing. The means and standard deviations of responses on the Guessing scale are reported in Appendix Tables M and N, and the results of the analysis of variance for that scale are given in Table 9.

Figure 14
Three-Way Interaction of Mode of Administration, Race, and Order of Administration for Guessing Scale Scores

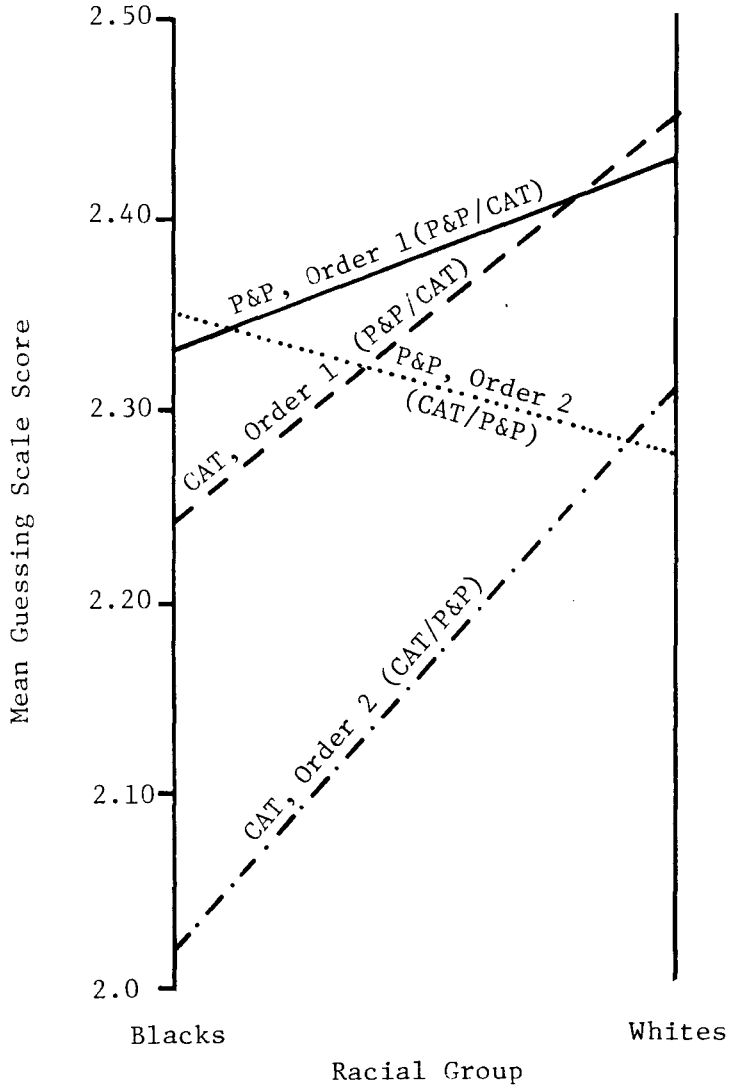


Table 9
Results of the Analysis of Variance for Guessing Scale Scores

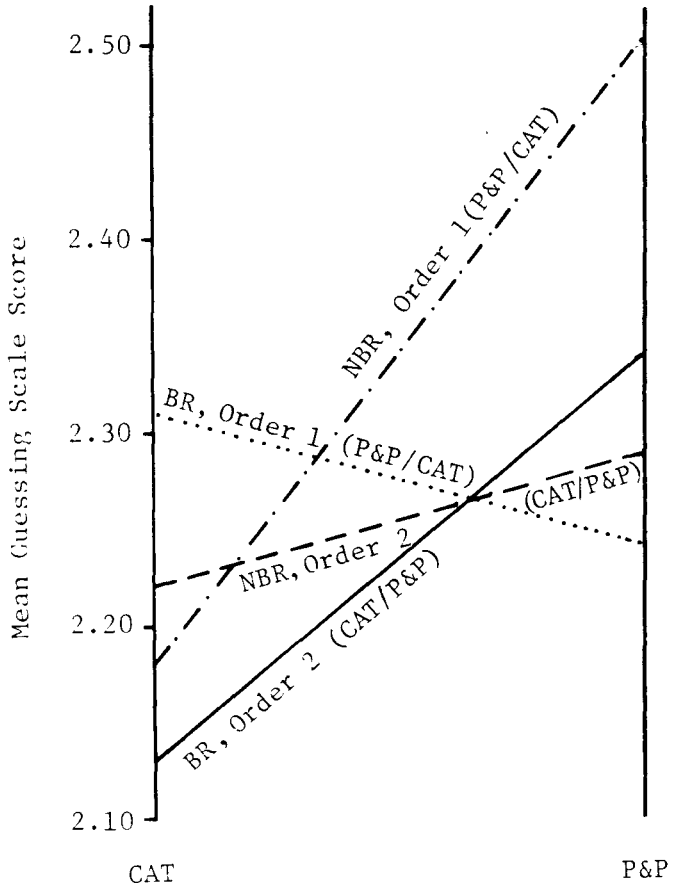
Source of Variation	Degrees of Freedom	Mean Square	F	p*
Between Subjects				
Main Effects				
Race	1	.47	.63	.429
Order	1	.51	.68	.412
Knowledge of Results (KR)	1	.10	.14	.710
Bias Reduction (BR)	1	.33	.44	.507
Two-Way Interactions				
Race × Order	1	.00	.00	.953
Race × KR	1	1.73	2.31	.130
Race × BR	1	.05	.06	.800
Order × KR	1	.21	.28	.596
Order × BR	1	.18	.24	.627
KR × BR	1	.25	.34	.562
Three-Way Interactions				
Race × Order × KR	1	1.34	1.79	.183
Race × Order × BR	1	.22	.29	.591
Race × KR × BR	1	.17	.22	.636
Order × KR × BR	1	2.74	3.67	.057
Four-Way Interaction				
Race × Order × KR × BR	1	.47	.63	.427
Error	185	.75		
Within Subjects				
Main Effect				
Mode	1	2.06	6.06	.015
Two-Way Interactions				
Mode × Race	1	.34	1.01	.316
Mode × Order	1	.00	.01	.936
Mode × KR	1	.04	.11	.739
Mode × BR	1	.85	2.52	.114
Three-Way Interactions				
Mode × Race × Order	1	1.26	3.72	.055
Mode × Race × KR	1	.17	.50	.480
Mode × Race × BR	1	.03	.08	.781
Mode × Order × KR	1	.27	.79	.375
Mode × Order × BR	1	1.53	4.51	.035
Mode × KR × BR	1	.01	.04	.838
Four-Way Interactions				
Mode × Race × Order × KR	1	.06	.19	.664
Mode × Race × Order × BR	1	.00	.00	.963
Mode × Race × KR × BR	1	.01	.04	.850
Mode × Order × KR × BR	1	.22	.65	.421
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.38	1.13	.290
Error	185	.34		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Again, there was a statistically significant ($p<.02$) main effect for mode of administration, with all students reporting that they guessed more often on the conventional paper-and-pencil tests (mean = 2.34) than on the computer-administered adaptive tests (mean = 2.21; see Table N). The interpretation of this difference was complicated by the significant Mode×Race×Order interaction ($p=.055$), shown in Figure 14. In three of the four Mode×Order conditions, the Black students reported that they guessed less than did the White students. Lowest levels of guessing were reported by Black students on the computer-administered adaptive tests, particularly when the computerized test was administered first (Order 2). White students reported highest levels of guessing, on both the adaptive and paper-and-pencil tests, when the paper-and-pencil test was administered first (Order 1); they reported lowest levels of guessing on both tests when the adaptive test was administered first (Order 2).

The three-way Mode×Order×BR interaction was also statistically significant ($p=.035$) and is shown in Figure 15. The highest level of guessing was reported on the NBR paper-and-pencil test when it was administered first (Order 1). Lowest levels of guessing were reported on the BR adaptive test

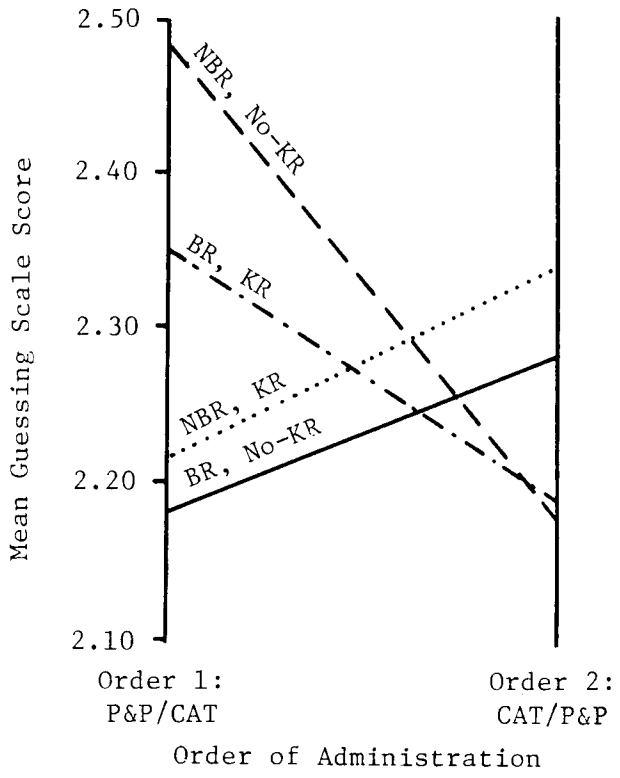
Figure 15
Three-Way Interaction of Mode of Administration, Order of Administration, and Bias-Reduction (BR) for Guessing Scale Scores



when it was administered first (Order 2). For three of the four comparisons between the adaptive and conventional tests, lower levels of guessing were reported on the adaptive tests; the exception was the BR adaptive test when it was administered first in the pair (Order 1).

The three-way Order×KR×BR interaction was also marginally significant (p=.057); Figure 16 shows the mean guessing scores for these test administration conditions. Highest levels of guessing were reported under Order 1 (P&P/CAT) when the NBR test was administered without KR; when the same test was administered under the reverse order, lowest levels of guessing were reported.

Figure 16
Three-Way Interaction of Order of Administration, Knowledge of Results (KR), and Bias-Reduction (BR) for Guessing Scale Scores



Relationship Between Ability Estimates and Psychological Reactions

Table 10 shows the Pearson product-moment correlations between the Bayesian ability estimates for the conventional paper-and-pencil and computerized adaptive tests and corresponding scores on the psychological reaction scales for each test. These data show that the only psychological variable which was not related to ability scores was reported motivation. There was a small to moderate tendency for students who performed better on the tests to be less nervous (r=-.25 for the paper-and-pencil test; r=-.16 for the adaptive test) and to report less tendency to guess (r=-.30) for the paper-and-pencil test. The strongest relationship was between ability

scores and students' reactions to knowledge of results. Higher ability students felt better about receiving KR (r=-.44 for the paper-and-pencil test; r=-.38 for the adaptive test) than lower ability students. This is not surprising in the paper-and-pencil test, where lower ability students would receive more negative feedback on their performance; but the effect also held for the adaptive test, which should have provided comparable amounts of positive and negative feedback for high- and low-ability students. In all cases where the psychological variables related to the ability scores (i.e., nervousness, reaction to knowledge of results, and guessing), the relationship between these variables was stronger in the conventional paper-and-pencil test than in the computerized adaptive test. This may indicate a "homogenizing" effect on students' reactions to testing when tests are administered adaptively by computer.

Table 10
Correlations of Bayesian Ability Estimates on the Conventional Paper-and-Pencil (P&P) and Computerized Adaptive Tests (CAT) with Psychological Reactions Scale Scores

Test	Nervousness	Knowledge of Results (KR)	Motivation	Guessing
P&P	-.25**	-.44**	.03	-.30**
CAT	-.16*	-.38**	-.04	-.05

*p<.05; **p<.01

DISCUSSION

The results indicate that the bias-reduced strategy of test construction used in this study to reduce racial performance differences was partially successful. Although the BR tests contained some items which clearly favored Black students, the majority of the items represented only a reduction in the degree to which the items favored White students over the NBR tests. In general, the White students obtained higher ability estimates than the Black students. However, mean ability estimates for the Black students were comparable to those of the White students on both the conventional paper-and-pencil and computerized adaptive tests when the bias-reduced tests were given without the provision of KR. When KR was provided on the BR tests, Black students obtained significantly lower mean ability estimates than White students.

This negative effect of KR appears to be contrary to the earlier reported data (Weiss, 1975) showing that KR itself eliminated mean racial differences in scores. What is similar between the two studies, however, is the finding that certain combinations of test administration conditions can reduce mean racial differences in ability estimates to nonsignificant levels. These results suggest that observed racial differences in verbal ability may be largely a function of test administration conditions, rather than a reflection of true racial differences.

The differences in the effects of KR on the Black students in this study and in the previous study may have been the result of differences in the way

KR was administered. In the earlier study the KR administered to both groups was designed to be specifically meaningful to the Black students. That is, KR was administered in terms which were derived from Black high school students, such as "right on." This form of feedback may have been more motivating to the Black students than the more typical feedback terms used in the present study. Black students in this study did report less favorable reactions to KR than White students, indicating that it "made them nervous" and "inhibited their concentration," thus potentially interfering with their test performance.

Another possible reason for the relatively high performance of Black students on the BR test under No-KR conditions and low performance under KR conditions relates to the item characteristics and difficulties of the tests. As mentioned above, the BR tests contained some words which were more appropriate for Black students, but the majority of the words represented only a reduction in the degree to which the items favored White students over the NBR test. Analysis of the nervousness reaction data indicated that the Black students were less nervous in the BR condition, presumably because some of the items appeared to be more appropriate for them. This effect was strongest for the paper-and-pencil test, as was the combined effect of bias-reduction and no knowledge of results for ability scores. While reduced nervousness may have aided performance on BR tests when No-KR was provided, BR performance was markedly reduced when KR was given, especially on the paper-and-pencil test. In the paper-and-pencil test this may have been due to the fact that the mean ability level for the Black students was lower than the ability level at which the conventional test was peaked. Thus, while the BR tests should have appeared to be easier for the Black students than the NBR tests, substantial negative feedback would have been received under the KR condition, possibly offsetting the positive psychological effects of taking the BR tests without receiving knowledge of results. When a Black student responded incorrectly to an item, in effect, the student was being told that he or she did not know the meaning of a "Black-type" word. It seems reasonable that negative feedback would have a stronger effect under these circumstances than in the NBR condition, an interpretation which is consistent with the result that Black students were less favorable to KR than White students.

This interpretation suggests that the motivational effects of KR may depend on the difficulty of the test for an examinee and, in particular, the proportion of negative versus positive feedback which the examinee receives. Figure 4 shows that for the Black students the negative effect of KR as provided in this study was stronger in the conventional paper-and-pencil test than in the computer-administered test. This may be due to the adaptive nature of the computer-administered test, which tends to equalize the amount of negative and positive KR each student receives, thus possibly reducing the adverse effects of negative KR.

The measurement properties of the BR tests were not as good as those of the NBR tests. Because of their item selection strategy, the NBR tests were substantially more discriminating than the BR tests. Related to this increased item discrimination was the increased precision of ability estimates in the NBR tests as indexed by the Bayesian posterior variances of these

estimates. The lower levels of discrimination in the BR tests are consistent with the finding of Church et al. (1978) that "Black-type" words are less discriminating than more standard vocabulary test words for both Black and White students.

The data also permit some conclusions regarding conventional and adaptive testing strategies. Correlations of ability estimates across the two testing modes found substantial (r=.73), but not perfect, agreement between individual ability estimates. The distributions of the two sets of estimates suggested that divergence from stronger agreement was in part due to the adaptive test spreading individuals out more on the ability continuum than did the peaked tests. This may reflect the better measurement in the tails of the distribution, which is typical of adaptive tests. More equi-precise measurement was apparent in this study when the computerized adaptive and conventional paper-and-pencil tests were both non-bias-reduced. Under this condition, the ability estimates from the computer-administered adaptive test had smaller posterior variances except in the range of abilities where the paper-and-pencil test was peaked. For the BR tests, the paper-and-pencil test was more precise except for low-ability students. This differential effectiveness of the adaptive test under BR and NBR conditions implies that the selection of items within strata in stradaptive tests should be on the basis of item discriminations if the desired result is maximum precision, as has been suggested by Weiss (1974).

The data also show (Figure 4 and Table 10) that computerized adaptive testing also reduced the effects of other variables (e.g., KR, BR) on mean ability test performance in comparison to conventional paper-and-pencil test administration.

The clearest findings from the present study relate to the psychological effects of adaptive and conventional tests and the KR and BR variables on the two racial groups. The computer-administered adaptive test motivated both racial groups more than the conventional paper-and-pencil tests, as reflected in the significant main effect for the motivation dependent variable. The significant Mode×Order×Race×BR interaction for motivation scores (see Figure 13) indicated that under both BR and NBR conditions, the motivation level of the Black students was much lower on the paper-and-pencil test when it was taken second (Order 2). With the exception of the NBR adaptive test in Order 1, which was the only condition free of order effects, the Black students reported higher levels of motivation on the computerized adaptive test as compared to the paper-and-pencil test under both BR and NBR conditions. The strong order effect observed for the Black students was not generally found for the White students. For White students, motivation was highest for the adaptive test except when it was bias-reduced and was taken second.

The generally higher levels of reported motivation on the computer-administered adaptive test for both groups, but especially for the Black students, may have been a joint function of the novel testing format and the adaptive nature of the test; the test should have appeared less difficult than the conventional paper-and-pencil test, which was peaked above both racial groups' mean ability levels. The fact that the computerized adaptive test was able to actually increase motivation when it was given second, in contrast to the apparent fatigue effect (especially for Black students) when the paper-and-

pencil test was given second, is especially encouraging for the use of this mode of test administration.

The data in Figure 13, and the marginally significant Mode×Order×Race effect which it subsumes, suggest that the motivation of Black students suffered more when the paper-and-pencil test was given second. The data also suggested that Black students preferred to take the paper-and-pencil test first and the computerized adaptive test second, while White students preferred the opposite. This significant Mode×Order×Race effect appeared elsewhere in the results. For the number of omitted responses variable, Black students omitted the most items when the paper-and-pencil test was taken second (see Figure 8) and the fewest items when the paper-and-pencil test was taken first. The opposite was true for the White students. Similarly for the guessing variable (see Figure 14), Black students reported guessing least (omitted more, were less motivated) when the paper-and-pencil test was administered second, while the White students guessed more when this test was administered first. These findings suggest that the differential sequential effect of the computer-administered and paper-and-pencil tests may be greater for the Black students. That is, once the novel computer-administered adaptive test had been given, the Black students seemed less interested in taking a conventional paper-and-pencil test. This would support the general conclusion of Johnson and Mihal (1973) that conditions of test administration can affect test-taking motivation.

Interestingly, while both Black and White students reported higher motivation on the computer-administered adaptive tests, they also reported more nervousness for this condition, as reflected in the significant main effect for the mode factor with the nervousness dependent variable. In fact, the significant Mode×Order interaction for nervousness (see Figure 10) indicated that when the computerized adaptive test was given first (Order 2), the increased nervousness carried over into the paper-and-pencil test, which was given second. When the paper-and-pencil test was given first (Order 1), nervousness was substantially lower until the computerized adaptive test was given, at which time it rose sharply. The higher reported motivation, but also nervousness, associated with administration of computerized adaptive tests suggests that during the administration of this test there was a general increased level of arousal or attention.

A further possible advantage of the computerized adaptive test over the conventional paper-and-pencil test was that students reported more guessing on the paper-and-pencil tests, which may be due to the fact that the adaptive test presented more items closer to the student's ability level. This apparent advantage resulted from the fact that the point at which the paper-and-pencil test had been peaked was above the ability level of the students. It is supported by the finding that higher ability students, besides reporting less nervousness, also reported less guessing.

A final interesting difference between the two modes of administration involves the differential relation of actual ability estimates to the various psychological reactions. Three of the four psychological dependent variables (reaction to knowledge of results, nervousness, and guessing) had statistically significant correlations with ability estimates in the expected direction. Thus, higher ability students reported more favorable reactions to knowledge

of results, less nervousness, and less guessing. In all three cases, the relationship between estimated ability levels and psychological reactions was stronger for the conventional paper-and-pencil test. This supports the important conclusion that the computer-administered adaptive test was successful in reducing the effects of extraneous variables on test performance and is consistent with the findings and interpretation above, which suggested that Black students were less tolerant of paper-and-pencil tests and that both groups were more motivated on the adaptive test.

The data also showed racial differences in reactions to the provision of knowledge of results. While Black students felt less favorable about KR, as indicated earlier, a significant Race×KR interaction for the number of omitted responses score indicated that the presence of KR induced Black students to omit fewer items than under the No-KR condition. White students omitted the same average number of responses under both conditions. Thus, while KR made Black students more nervous, it also caused them to omit fewer responses. This implies that similar to the effects suggested above for computerized administration, the KR condition caused an increase in general arousal, or interest in one's performance. While this arousal could take the form of nervousness, reaction to KR was more favorable for both groups during the second test (see Figure 9), suggesting a familiarity effect.

Conclusions

Selection of items on the basis of an index of bias has been shown to reduce racial differences in mean performance on verbal ability tests when other variables, such as motivational factors, do not interfere with the effect. Since item selection based on bias-reduction alone can result in less precise measurement, simultaneous consideration of more traditional item statistics, such as item discrimination, should also be made in the development of bias-free tests.

The differential motivational impact of computer-administered versus conventional paper-and-pencil tests was given strong support in this study, and there were several indications that the psychological contrast between computer-administered and paper-and-pencil tests may differ for Black and White students. If this can be replicated, it may be possible to obtain more comparable motivational states across racial groups using computer-administered tests. In addition, the reaction to provision of knowledge of results differed for Black and White students.

This study has shown that ability test scores, and the reactions of different groups to ability tests, are to some extent a function of the conditions under which these tests are administered. The results support earlier studies on the effects of test administration conditions on both ability test scores and psychological reactions to testing (e.g., Betz & Weiss, 1976; Prestwood & Weiss, 1978). These data imply the need for further study of the effects of test administration conditions on members of minority groups to determine those administration conditions which maximize their ability estimates either directly or through their effects on the psychological environment of testing.

REFERENCES

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude (Research Bulletin RE 71-59). Princeton, NJ: Educational Testing Service, October 1971, pp. 1-24.

Betz, N. E. Prospects: New types of information and psychological implications. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. AD A027170).

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1977. (NTIS No. AD A046062).

Church, A. T., Pine, S. M., & Weiss, D. J. A comparison of levels and dimensions of performance in Black and White groups on tests of vocabulary, mathematics, and spatial ability (Research Report 78-3) Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978. (NTIS No. AD A062797).

DeWitt, L. J., & Weiss, D. J. A computer software system for adaptive ability measurement (Research Report 74-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD 773961).

Johnson, D. I., & Mihal, W. M. Performance of Blacks and Whites in computerized versus manual testing environments. American Psychologist, 1973, 28, 694-699.

Larkin, K. C., & Weiss, D. J. An empirical comparison of two-stage pyramidal adaptive ability testing (Research Report 75-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A027147).

Martin, J. T., Pine, S. M., & Weiss, D. J. An item bias investigation of a standardized aptitude test (Research Report 78-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement (Research Report 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD 781894).

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. AD A022964).

Owen, R. A Bayesian sequential procedure for quantal response in a context of adaptive verbal testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Pine, S. M., & Weiss, D. J. A comparison of the fairness of adaptive and conventional testing strategies (Research Report 78-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978. (NTIS No. AD A059436).

Prestwood, J. S., & Weiss, D. J. The effects of knowledge of results and test difficulty on ability test performance and psychological reactions to testing (Research Report 78-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

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

Vale, C. D. Strategies of branching through an item pool. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD 768316).

Weiss, D. J. Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD A004270).

Weiss, D. J. (Ed.). Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

APPENDIX: SUPPLEMENTARY TABLES

Table A
Examples of Vocabulary Items

Items from Black Literature and Black Psychologist

Ranking	Gatemouth
1. Murdering	1. Gossiper
2. Exchange of insults	2. Doorway
3. Pig's intestines	3. Jazz musician
4. Fried cow's tail	4. Dog
5. Olympic event	5. Fat person
Shiv	**Swag**
1. Politician	1. Construction worker
2. Genius	2. Beggar
3. Book	3. Corrupt politician
4. Drifter	4. Stolen goods
5. Knife	5. Garbage

"White" type items from Webster's Seventh Collegiate Dictionary

Borsch	Torte
1. Overcoat	1. Cake
2. Dog	2. Twist
3. Porter	3. Shirt
4. Soup	4. Crime
5. Chamber	5. Answer
Afghan	**Gefilte Fish**
1. Alien	1. Type of fish
2. Harbor	2. A game
3. Canvas	3. Food
4. Vista	4. A sport
5. Blanket	5. Sucker

Items from Standardized Vocabulary Tests

Accumulate	Reinforce
1. Become cloudy	1. Speak loudly
2. Get angry	2. Come again to
3. Get dirty	3. Revise
4. Imitations	4. Apply again
5. Claws	5. Make stronger
Oppressed	**Capitulate**
1. Wrinkled	1. Entitle
2. Expressed	2. Surrender
3. Musically talented	3. Behead
4. Disowned	4. Put in charge
5. Put down	5. Congratulate

Table B

Item Numbers, Discrimination (a), Difficulty (b), and Bias Parameters for Items in the Vocabulary Stradaptive Item Pool (c=.20 for All Items)

Item	a	b	Bias	Item	a	b	Bias	Item	a	b	Bias	Item	a	b	Bias
Stratum 7:				Stratum 5, *cont'd.*:				Stratum 4, *cont'd.*:				Stratum 3, *cont'd.*:			
(15 items)				1300	1.22	.71	.64	1315	1.16	-.05	.95	1281	.84	-.38	1.40
152	2.87	1.70	.35	1227	1.19	.34	.90	1429	1.11	.24	1.26	1408	.82	-.31	1.11
162	1.58	1.57	-.18	1290	1.18	.34	.72	628	1.09	-.22	.76	1412	.81	-.60	1.01
166	1.19	2.07	.08	1425	1.17	.68	.85	1289	1.08	.17	1.39	202	.81	-.47	.88
1256	1.12	1.52	.96	1251	1.11	.84	2.00	1228	1.07	.26	.19	1287	.73	-.47	1.05
114	1.05	2.34	.06	1219	1.10	.38	.41	191	1.05	-.02	.24	88	.62	-.59	1.30
1276	1.03	1.60	1.02	1244	1.06	.39	.51	1423	1.02	.01	.60	1248	.35	-.50	-1.11
1229	1.01	2.40	-1.20	127	1.05	.55	.52	1298	1.01	-.28	.60	1301	.28	-.53	-.21
1220	.90	1.90	.26	1265	1.03	.32	1.55	1284	1.01	-.09	.94				
1249	.64	2.31	1.53	1213	.98	.47	1.12	501	1.00	.03	.27	Stratum 2:			
1264	.58	2.54	.52	1321	.96	.68	.52	1260	.99	.24	-.12	(16 items)			
1206	.28	1.54	.09	1325	.87	.69	1.09	1313	.89	-.23	.47	1407	1.78	-1.42	.59
1232	.26	3.18	-3.36	1225	.84	.48	.40	1268	.85	-.00	.11	65	1.29	-1.06	.54
1245	.22	3.98	-4.44	1238	.79	.64	2.97	1231	.85	.25	.77	1409	1.25	-1.15	.92
1257	.22	1.61	-3.08	1297	.76	.45	1.04	1324	.84	-.10	.76	1259	1.20	-.91	.30
1278	.22	3.43	-3.97	1280	.75	.34	.16	1236	.76	-.26	.51	182	1.14	-1.47	.70
				95	.70	.40	1.40	1208	.74	.22	.92	25	1.12	-1.30	.96
Stratum 6:				648	.59	.87	.66	63	.69	.11	.30	1405	1.09	-1.04	.59
(17 items)				237	.48	.88	-.32	1316	.65	.25	1.46	1282	1.08	-.91	.79
1291	2.08	1.02	1.39	1204	.43	.52	.41	1252	.56	-.11	-.54	1235	.79	-.94	.01
1303	1.52	1.14	1.07	322	.41	.82	1.33	1299	.53	-.03	-.20	1311	.77	-1.22	1.35
43	1.49	1.00	1.57	1202	.33	.65	-.10	199	.52	-.04	.54	1262	.76	-.94	-.35
1274	1.44	1.21	1.28	1242	.09	.80	-4.57	101	2.07	.16	.62	1322	.74	-.96	1.02
1243	1.22	1.23	1.28	1223	.09	.84	-5.46					19	.73	-1.20	.73
1317	1.19	1.26	1.03	1216	1.01	.57	.59	Stratum 3.:				1312	.62	-.91	.82
188	1.18	.95	1.01	173	1.68	.33	.99	(25 items)				1318	.60	-1.05	1.06
1201	.98	1.01	.34	235	1.38	.86	1.03	96	1.79	-.42	.74	1406	.53	-.92	.08
1240	.97	.91	.74	51	1.76	.68	1.09	23	1.42	-.45	1.18				
1222	.90	1.29	1.37					1320	1.31	-.62	.60	Stratum 1:			
1217	.74	1.27	.88	Stratum 4:				1323	1.20	-.40	.68	(11 items)			
1215	.72	1.26	1.35	(35 items)				86	1.16	-.36	.88	1400	1.49	-1.64	.26
1279	.68	1.05	1.61	27	2.10	-.27	.59	28	1.12	-.74	1.00	1404	1.32	-1.57	.44
1438	.64	1.36	.22	123	2.03	.07	.25	1293	1.07	-.47	.64	1402	1.13	-1.79	.58
1309	.48	1.10	-.30	1410	1.79	-.14	.98	1414	1.04	-.73	.70	122	1.10	-1.91	.77
1305	1.33	1.14	1.03	1419	1.54	-.13	.75	1403	1.04	-.89	1.09	1234	.90	-1.83	.66
52	1.70	1.46	.90	1422	1.50	.01	1.04	32	.96	-.59	.94	71	.75	-1.94	2.05
				190	1.46	-.24	.34	1212	.95	-.87	1.33	66	.69	-1.64	-.20
Stratum 5:				1426	1.36	-.09	.91	1304	.94	-.31	.77	1272	.63	-1.59	.30
(32 items)				1207	1.34	.16	.44	1246	.94	-.88	.62	1275	.56	-2.04	-.03
112	2.22	.32	.86	47	1.32	.27	-.20	1421	.93	-.42	.23	1239	.52	-1.66	-.36
106	1.61	.51	.63	1420	1.32	.01	1.20	1295	.92	-.77	1.08	240	.41	-1.95	3.73
1430	1.27	.50	1.16	5	1.25	-.26	.80	1310	.90	.85	.34				
116	1.26	.88	.89	1413	1.21	.25	1.11	16	.90	-.80	1.25				

Note. For the bias-reduced adaptive test, the items were ordered within each stratum in increasing order of bias.

Table C
Item Numbers, Discrimination (a), Difficulty (b), and Bias Parameters for Items in the Vocabulary Pencil-and-Paper Tests

Non-Bias-Reduced				Bias-Reduced			
Item No.	a	b	Bias	Item No.	a	b	Bias
52	1.70	1.46	.90	1204	.43	.52	.41
1302	1.50	.87	1.21	1414	1.04	.73	.70
1418	1.52	.13	.84	501	1.00	.03	.27
189	1.78	.60	1.71	1211	2.27	.18	.28
85	1.54	.29	.39	1223	.09	.84	-5.46
22	1.42	- .13	.90	1260	.99	.24	- .12
311	1.50	.12	.22	47	1.32	.27	- .20
1305	1.33	1.14	1.03	1240	.97	.91	.74
105	1.37	- .40	1.12	181	1.17	- .66	.65
1401	1.59	-1.48	.37	1401	1.59	-1.48	.37
1411	1.85	.09	1.24	191	1.05	- .02	.24
1254	1.68	- .60	.95	1323	1.20	- .40	.68
285	1.48	.35	.74	1201	.98	1.01	.34
1415	1.29	- .69	.76	1219	1.10	.38	.41
1416	2.04	- .21	.74	1228	1.07	.26	.19
1211	2.27	.18	.28	1244	1.06	.39	.51
51	1.76	.68	1.09	311	1.50	.12	.22
522	1.36	.12	.62	1299	.53	- .03	- .20
121	1.18	- .92	.89	1235	.79	- .94	.01
181	1.17	- .66	.65	1248	.35	- .50	-1.11

Table D
Test Reaction Questions, by Scale

Scaled Score	Knowledge of Results Scale
	DID RECEIVING FEEDBACK AFTER EACH QUESTION INTERFERE WITH YOUR ABILITY TO CONCENTRATE ON THE TEST?
1	NO, NOT AT ALL
2	YES, SOMEWHAT
3	YES, MODERATELY SO
4	YES, VERY MUCH SO
	DID GETTING FEEDBACK AFTER EACH QUESTION MAKE YOU NERVOUS?
1	NO, NOT AT ALL
2	YES, SOMEWHAT
3	YES, MODERATELY SO
4	YES, VERY MUCH SO

Scaled Score	Nervousness Scale
	WERE YOU NERVOUS WHILE TAKING THE TEST?
1	NOT AT ALL
2	SOMEWHAT
3	MODERATELY SO
4	VERY MUCH SO
	DID NERVOUSNESS WHILE TAKING THE TEST PREVENT YOU FROM DOING YOUR BEST?
4	YES, DEFINITELY
3	YES, SOMEWHAT
2	PROBABLY NOT
1	DEFINITELY NOT

Scaled Score	Motivation Scale
	DID YOU CARE HOW WELL YOU DID ON THE TEST?
4	I CARED A LOT
3.2	I CARED SOME
2.4	I CARED A LITTLE
1.6	I CARED VERY LITTLE
.8	I DIDN'T CARE AT ALL

Scaled Score	Motivation Scale (cont'd.)
	DID YOU FEEL CHALLENGED TO DO AS WELL AS YOU COULD ON THE TEST?
1	NOT AT ALL
2	SOMEWHAT
3	FAIRLY MUCH SO
4	VERY MUCH SO
	WERE YOU INTERESTED IN KNOWING WHETHER YOUR ANSWERS WERE RIGHT OR WRONG?
4	I WAS VERY INTERESTED
3	I WAS MODERATELY INTERESTED
2	I WAS SOMEWHAT INTERESTED
1	I DIDN'T CARE AT ALL

Scaled Score	Guessing Scale
	ON HOW MANY OF THE QUESTIONS DID YOU GUESS?
4	ALMOST ALL OF THE QUESTIONS
3.33	MORE THAN HALF OF THE QUESTIONS
2.67	ABOUT HALF OF THE QUESTIONS
2	LESS THAN HALF OF THE QUESTIONS
1.33	ALMOST NONE OF THE QUESTIONS
.67	NONE OF THE QUESTIONS
	HOW OFTEN WERE YOU SURE THAT YOUR ANSWERS TO THE QUESTIONS WERE CORRECT?
.8	1. ALMOST ALWAYS
1.6	2. MORE THAN HALF OF THE TIME
2.4	3. ABOUT HALF OF THE TIME
3.2	4. LESS THAN HALF OF THE TIME
4	5. ALMOST NEVER

Table E
Means and Standard Deviations of Bayesian Posterior Ability Estimates as a Function of Ability Estimates for Adaptive and Conventional Tests in Non-Bias-Reduced and Bias-Reduced Conditions

Bayesian Ability Estimate Interval		Non-Bias-Reduced Condition						Bias-Reduced Condition					
		Conventional Test			Adaptive Test			Conventional Test			Adaptive Test		
Lo	Hi	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
-2.0	-1.81	0	--	--	7	.080	.006	2	.169	.002	6	.114	.016
-1.8	-1.61	2	.103	.000	2	.089	.012	8	.195	.017	4	.131	.012
-1.6	-1.41	7	.116	.013	6	.092	.026	12	.180	.035	6	.122	.028
-1.4	-1.21	19	.114	.016	8	.078	.008	10	.158	.037	6	.142	.033
-1.2	-1.01	14	.115	.018	9	.077	.009	10	.189	.031	9	.134	.013
-1.0	- .81	20	.106	.025	13	.092	.024	13	.174	.043	8	.177	.054
- .8	- .61	12	.098	.023	16	.082	.008	11	.143	.018	6	.152	.033
- .6	- .41	8	.089	.021	15	.103	.034	4	.130	.012	9	.160	.031
- .4	- .21	8	.088	.015	9	.106	.037	12	.150	.014	15	.180	.027
- .2	- .01	11	.073	.046	7	.093	.007	15	.114	.062	10	.136	.100
.0	.19	4	.086	.008	10	.097	.028	3	.132	.016	8	.184	.029
.2	.39	4	.079	.003	3	.082	.017	4	.128	.005	7	.227	.054
.4	.59	2	.087	.008	1	.068	.000	1	.140	.000	5	.233	.024
.6	.79	0	--	--	0	--	--	3	.139	.008	0	--	--
.8	.99	0	--	--	0	--	--	2	.155	.005	2	.289	.028
1.0	1.19	2	.139	.009	0	--	--	0	--	--	0	--	--
1.2	1.39	0	--	--	1	.134	.000	0	--	--	0	--	--
1.4	1.59	0	--	--	0	--	--	0	--	--	0	--	--
1.6	1.79	0	--	--	0	--	--	0	--	--	0	--	--
1.8	2.00	0	--	--	1	.209	.000	0	--	--	1	.204	.000

Table F
Means and Standard Deviations of Bayesian Ability Estimates
for all Combinations of the Independent Variables

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	-.94	-.86	-.59	-.90	-.88	-.81	-1.13	-.82
S.D.	1.02	.95	1.07	.93	.84	.94	.57	.58
P&P								
N	15	14	13	14	15	13	12	14
Mean	-1.05	-.98	-.51	-.73	-.79	-.99	-.91	-.84
S.D.	.66	.77	.81	.62	.48	.71	.50	.62
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	-.75	-.32	-.77	-.56	-.58	-.60	-.72	-.63
S.D.	.67	.78	.66	.88	.64	.74	.75	.70
P&P								
N	13	14	11	13	14	15	15	13
Mean	-.66	-.40	-1.04	-.58	-.78	-.68	-.42	-.55
S.D.	.68	.54	.51	.62	.45	.50	.67	.57

Table G
Means and Standard Deviations of Bayesian Posterior Variances
for all Combinations of the Independent Variables

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	.15	.17	.16	.16	.09	.11	.09	.08
S.D.	.04	.04	.04	.04	.02	.04	.01	.02
P&P								
N	15	14	13	14	15	13	12	14
Mean	.18	.16	.16	.15	.10	.11	.10	.10
S.D.	.04	.04	.03	.03	.03	.02	.02	.01
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	.15	.19	.15	.20	.11	.09	.09	.08
S.D.	.03	.06	.03	.06	.04	.02	.02	.01
P&P								
N	13	14	11	13	14	15	15	13
Mean	.17	.14	.17	.15	.11	.09	.10	.10
S.D.	.04	.03	.03	.04	.02	.02	.03	.03

Table H
Means and Standard Deviations of Number of Omitted Responses
Under All Combinations of the Independent Variables

Group and Mode	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	3.00	1.38	3.08	3.38	2.13	1.23	4.33	4.50
S.D.	3.57	2.02	2.93	3.50	2.62	1.92	4.10	2.98
P&P								
N	15	14	13	14	15	13	12	14
Mean	2.07	2.86	1.77	4.50	1.87	.23	4.42	5.00
S.D.	4.50	4.79	2.89	6.21	2.82	.60	5.45	6.67
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	3.17	2.27	2.73	3.00	2.21	1.73	3.33	3.60
S.D.	4.49	2.28	2.80	2.42	2.72	1.98	4.62	3.18
P&P								
N	13	14	11	13	14	15	15	13
Mean	3.46	1.93	2.82	2.46	5.21	2.27	2.53	2.85
S.D.	5.08	2.76	3.84	3.50	6.23	4.25	4.27	4.26

Table I
Means and Standard Deviations of Dependent Variables for the Combined Racial, Bias Reduction, Knowledge of Results, Order of Administration, and Mode of Administration Groups

	Bayesian Scores			Posterior Variance			Number of Omits		
Combined Groups	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
Racial									
Blacks									
CAT	108	-.87	.86	108	.12	.05	108	2.87	3.14
P&P	110	-.85	.66	110	.13	.04	110	2.83	4.75
Whites									
CAT	112	-.61	.72	112	.13	.06	112	2.73	3.12
P&P	108	-.63	.58	108	.13	.04	108	2.94	4.36
Bias Reduction									
Bias-Reduced									
CAT	107	-.70	.88	107	.17	.05	107	2.72	3.03
P&P	107	-.74	.68	107	.16	.03	107	2.73	4.31
Non-Bias-Reduced									
CAT	113	-.76	.73	113	.09	.03	113	2.87	3.24
P&P	111	-.74	.58	111	.10	.02	111	3.03	4.80
Knowledge of Results									
Knowledge of Results									
CAT	112	-.71	.83	112	.13	.05	112	2.14	2.78
P&P	113	-.79	.62	113	.13	.04	113	2.49	4.29
No Knowledge of Results									
CAT	108	-.76	.78	108	.13	.05	108	3.47	3.34
P&P	105	-.69	.64	105	.13	.04	105	3.30	4.81
Order of Administration									
P&P/CAT									
CAT	107	-.79	.80	107	.12	.04	107	2.97	3.50
P&P	108	-.76	.63	108	.14	.04	108	2.98	4.52
CAT/P&P									
CAT	112	-.68	.81	112	.14	.06	112	2.65	2.74
P&P	110	-.72	.63	110	.13	.04	110	2.78	4.60
Mode of Administration									
CAT	220	-.73	.80	220	.13	.05	221	2.78	3.12
P&P	218	-.74	.63	218	.13	.04	218	2.88	4.55

Table J
Means and Standard Deviations of
Knowledge of Results Scale Scores

Group and Mode	Bias-Reduced Order 1 P&P/CAT	Bias-Reduced Order 2 CAT/P&P	Non-Bias-Reduced Order 1 P&P/CAT	Non-Bias-Reduced Order 2 CAT/P&P
Blacks				
CAT				
N	14	13	13	12
Mean	1.89	1.73	1.81	2.08
S.D.	.90	.90	.80	1.18
P&P				
N	13	14	15	13
Mean	2.31	1.64	1.83	1.69
S.D.	1.03	.72	.79	.75
Whites				
CAT				
N	12	13	12	14
Mean	1.38	1.58	1.38	1.54
S.D.	.43	.61	.53	.54
P&P				
N	11	14	13	12
Mean	1.46	1.29	1.31	1.46
S.D.	.47	.47	.44	.54

Table K
Means and Standard Deviations of the Nervousness Scale Scores

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	2.46	1.85	2.00	1.75	1.85	2.29	2.25	1.93
S.D.	.91	.69	.54	.51	.75	.72	.72	.80
P&P								
N	15	14	13	14	15	13	12	14
Mean	1.93	1.89	1.73	1.57	1.90	2.27	2.21	1.96
S.D.	.94	1.08	.70	.76	.71	.75	.78	.80
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.08	2.00	2.18	1.83	1.96	1.93	1.71	2.23
S.D.	.63	.68	.46	.62	.66	.62	.47	.56
P&P								
N	12	14	11	13	14	14	13	13
Mean	1.88	1.96	1.86	1.73	1.75	2.14	1.58	2.15
S.D.	.64	.66	.50	.60	.55	.82	.70	.56

Table L
Means and Standard Deviations of the Motivation Scale Scores

Group and Mode	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	3.52	2.66	2.90	3.14	3.02	3.27	3.17	2.83
S.D.	.56	.92	.49	.81	.47	.45	.48	.74
P&P								
N	15	14	13	14	15	13	12	14
Mean	3.37	2.34	2.90	3.17	2.96	2.62	3.02	2.66
S.D.	.64	.88	.61	.81	.66	.52	.60	.70
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.62	3.07	2.63	2.98	2.87	2.86	3.12	3.15
S.D.	.77	.71	.51	.56	.88	.82	.94	.59
P&P								
N	12	14	11	13	14	12	13	13
Mean	2.77	3.01	2.92	2.75	2.68	2.69	2.78	3.03
S.D.	.98	.52	.83	.56	.91	.76	.95	.48

Table M
Means and Standard Deviations of the Guessing Scale Scores

Group and Mode	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	2.44	2.08	2.15	1.87	2.10	2.24	2.24	1.97
S.D.	.78	.52	.59	.86	.76	.54	.70	.67
P&P								
N	15	14	13	14	15	13	12	14
Mean	2.32	2.37	2.07	2.33	2.20	2.58	2.77	2.12
S.D.	.86	.71	.66	.96	.80	.72	.58	.69
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.35	2.12	2.27	2.48	2.07	2.30	2.29	2.37
S.D.	.66	.72	.69	.49	.53	.62	1.02	.69
P&P								
N	12	14	11	13	14	12	13	13
Mean	2.30	2.19	2.27	2.48	2.48	2.18	2.63	2.27
S.D.	.93	.53	.97	.53	.79	.87	.87	.48

Table N
Means and Standard Deviations of the Test Reaction Scale Scores for the Combined Racial, Bias Reduction, Knowledge of Results, Order of Administration, and Mode of Administration Groups

	Knowledge of Results			Nervousness			Motivation			Guessing		
Combined Groups	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
Racial												
Blacks												
CAT	52	1.88	.93	105	2.04	.73	105	3.06	.67	105	2.14	.69
P&P	57	1.85	.84	110	1.93	.83	110	2.88	.74	110	2.33	.77
Whites												
CAT	51	1.47	.52	101	1.98	.59	101	2.92	.74	101	2.28	.69
P&P	50	1.37	.47	104	1.88	.65	102	2.83	.75	102	2.35	.77
Bias Reduction												
Bias-Reduced												
CAT	52	1.65	.75	102	2.02	.66	102	2.96	.72	102	2.22	.68
P&P	54	1.67	.79	106	1.82	.75	106	2.91	.77	106	2.29	.78
Non-Bias-Reduced												
CAT	51	1.70	.82	104	2.01	.68	104	3.03	.70	104	2.20	.70
P&P	53	1.58	.67	108	1.99	.73	106	2.81	.71	106	2.40	.75
Knowledge of Results												
Knowledge of Results												
CAT	103	1.68	.78	103	2.05	.72	103	2.99	.75	103	2.22	.64
P&P	105	1.63	.73	111	1.96	.78	109	2.82	.78	109	2.33	.79
No Knowledge of Results												
CAT				103	1.98	.61	103	3.00	.67	103	2.20	.74
P&P				103	1.84	.70	103	2.90	.70	103	2.36	.75
Order of Administration												
P&P/CAT												
CAT	51	1.63	.73	101	2.06	.68	101	3.00	.70	101	2.24	.72
P&P	53	1.74	.81	105	1.85	.71	105	2.93	.78	105	2.38	.82
CAT/P&P												
CAT	52	1.72	.84	105	1.97	.66	105	2.99	.72	105	2.17	.66
P&P	54	1.51	.63	109	1.96	.78	107	2.79	.70	107	2.31	.72
Mode of Administration												
CAT	103	1.68	.78	206	2.02	.67	206	2.99	.71	206	2.21	.69
P&P	107	1.63	.73	214	1.91	.74	212	2.86	.74	212	2.34	.77

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-249)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION & TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Head, Research, Development, and Studies (OP102X)
Office of the Chief of Naval Operations
Washington, DC 20370

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH CANTER
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

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

1 MCDEC
Quantico Marine Corps Base
Quantico, VA 22134

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

CoastGuard

1 MR. JOSEPH J. COWAN, CHIEF
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Vern W. Urry
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Warner Birice
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Progr
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Psychology
University of Illinois
Champaign, IL 61820

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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 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 982
JAPAN

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

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

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

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

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

PREVIOUS PUBLICATIONS

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-2. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

RELATIONSHIPS AMONG ACHIEVEMENT LEVEL ESTIMATES FROM THREE ITEM CHARACTERISTIC CURVE SCORING METHODS

G. Gage Kingsbury
and
David J. Weiss

RESEARCH REPORT 79-3
APRIL 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Office of Naval Research, Army Research Institute, and Air Force Human Resources Laboratory, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Relationships Among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. Gage Kingsbury and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.: RR042-04 T.A.: RR042-04-01 W.U.: NR150-389
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE April 1979 13. NUMBER OF PAGES 28
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 in 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

This research was supported by funds from the Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Army Research Institute, Air Force Human Resources Laboratory, and Office of Naval Research, and monitored by the Office of Naval Research.

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

latent trait test theory	achievement testing	testing
item response theory	computerized testing	tailored testing
response-contingent testing	adaptive testing	programmed testing
individualized testing	sequential testing	automated testing
	branched testing	

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

This study compared achievement level estimates from three item characteristic curve (ICC) scoring methods using the one-, two-, and three-parameter ICC models. The three scoring methods were maximum-likelihood normal, maximum-likelihood logistic, and Owen's (1975) Bayesian scoring method. Data included all possible response patterns from a hypothetical five-item test, as well as response patterns from live administration of a conventional and an adaptive achievement test. For the conventional and adaptive test data, correlations among achievement level estimates were examined as a function of test length.

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Results for all data sets showed a high degree of similarity among θ estimates for the one- and two-parameter data, with slight decreases in correlations as information on the discrimination parameter was used in scoring. When the third ("guessing") parameter was used in scoring the item response data, correlations among θ estimates were reduced, particularly for the adaptive test data. The data also showed an increasing tendency for the maximum-likelihood methods to result in convergence failures as the third parameter of the ICC was used in scoring. In general, however, the adaptive test data were less likely to result in convergence failures than were the conventional test data. The data also illustrated how each of the three scoring methods tend to utilize ICC parameter information in arriving at θ estimates and the relationships of these estimates to a number-correct scoring philosophy. Advantages and disadvantages of each of the scoring methods are discussed. It is suggested that future research examine the relative validities of scoring methods and model combinations.

CONTENTS

Acknowledgements

Adaptive and conventional test data utilized in this report were obtained from volunteer students in General Biology, Biology 1-011, at the University of Minnesota during fall quarter 1976 and winter quarter 1977; appreciation is extended to these students for their participation in this research. The cooperation of Kathy Swart and Norman Kerr of the General Biology staff in providing access to the students is also deeply appreciated.

Technical Editor: Barbara Leslie Camm

Relationships Among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods

With the advent of computerized instruction and testing, and the concurrent reduction in costs of minicomputer systems, it has become feasible to use item characteristic curve (ICC) response models to estimate students' achievement levels, based on responses to classroom tests. This feasibility has been demonstrated recently in an experimental context (Bejar, Weiss, & Kingsbury, 1977; Reckase, 1977), and computer programs for implementing these scoring methods have been made available (Bejar & Weiss, 1979). These technological advances should be paced by theoretical advances if perspective is to be maintained and the maximum possible return from advancing technology is to be insured.

When ICC response models are employed within a classroom situation, estimates of the achievement level of any student may be obtained in a number of different ways (Bejar & Weiss, 1979). The two most widely used scoring methods are the maximum-likelihood (M-L) estimators (Lord & Novick, 1968) and the Bayesian estimators (Lindgren, 1976). Estimates obtained by a M-L procedure will be asymptotically consistent and unbiased. The property of consistency implies that as the number of items answered by an individual increases toward infinity, the difference between the M-L estimate of a student's achievement level ($\hat{\theta}$) and the actual value of the parameter (θ) will approach zero. Therefore, as a test becomes very long, an estimate of the achievement level will approach the actual achievement level. The property of unbiasedness implies that if several M-L estimates of an achievement level are made, the mean of the estimates will equal the actual θ value. These properties are highly desirable from a statistical point of view.

Although estimates obtained using a Bayesian procedure (e.g., Owen, 1975) allow the incorporation of prior information into the achievement level estimation process, they are somewhat biased. This bias in the Bayesian achievement level estimates has been demonstrated by McBride and Weiss (1976) in a series of computer simulations. In each case the Bayesian scoring method was shown to provide θ estimates with average values different from the true θ levels that gave rise to the response pattern. Thus, individuals with a high true achievement level received an ability estimate that was lower than the true θ value, and individuals with a low true θ level received a θ estimate somewhat higher than the true value. The bias increased as the estimated θ became more discrepant from the true θ level.

Both M-L and Bayesian scoring methods allow the use of all of the information contained in the testee's responses to all the items in the test in order to arrive at the final estimate of the testee's achievement level. However, the Bayesian algorithm devised by Owen (1975) is somewhat affected by the order of the items in the test; that is, scoring the responses in a different order will result in a different estimate of trait level (Sympson,

1977). On the other hand, the M-L estimators are independent of the item order. In general, in a test of finite length, a single response pattern may receive differing achievement level estimates solely as a function of the differences between the scoring methods.

Samejima (1969) has noted that M-L estimates for individuals will differ as a function of the underlying response model. More importantly, though, she has pointed out that ordering of individuals' trait level estimates will change as a function of the response model assumed in the scoring method. Bejar and Weiss (1979) have also noted, within a two-parameter ICC model, that a difference in the ICC scoring method used will result in different trait level estimates for the same pattern of responses to the same test items. These investigators used all possible response patterns in a hypothetical five-item test to illustrate differences among three different methods for estimating trait levels; however, there is some question whether the differences found within the hypothetical data set used will generalize to live-testing data sets. According to ICC response theory, not all response vectors are equally likely. Because the hypothetical data sets used in the Samejima (1969) study and the Bejar and Weiss (1979) study were highly improbable--each possible response pattern occurred once--results from real data sets may reflect different levels of similarity among the results of different ICC scoring methods.

If differences in ordering of individuals as a function of the ICC scoring method are found in real data sets, such results will have direct consequences for educators who are preparing to implement a testing system utilizing ICC theory and procedures. In an educational situation, the ordering of individuals according to their responses on tests is of paramount importance. For this reason, it is important to determine the degree of disparity in achievement level estimates based on the different methods of scoring item responses using ICC theory. Similarly, since test response patterns can be scored by using one, two, or three of the parameters describing the ICC, different levels of similarity among θ estimates may be obtained by different scoring methods using each of the models.

The recent experimental applications of adaptive testing strategies in educational settings (e.g., Bejar, Weiss, & Gialluca, 1977; Brown & Weiss, 1977) may open the way to the use of shorter, more precise individualized tests in future classrooms. Since the Bejar and Weiss (1979) and Samejima (1969) data suggest that short tests may result in differences among achievement levels estimated by different scoring methods, it is imperative that the implementation of adaptive testing systems be accompanied by a knowledge of the differences among the achievement level estimates resulting from different scoring strategies for adaptively administered achievement tests. A beginning toward the development of this knowledge is simply the recognition that differences do exist among the various scoring methods and that these differences may have an impact on rankings of the individual students in the classroom. The present study was designed to investigate these differences through additional analyses of the data reported by Bejar and Weiss (1979) and Samejima (1969) and through analysis of data from the administration of conventional and adaptive tests.

Method

The three scoring methods described by Bejar and Weiss (1979) were compared across three different ICC response models. The three scoring methods were (1) maximum likelihood using a normal probability function (M-L normal), (2) maximum likelihood using a logistic probability function (M-L logistic), and (3) Owen's Bayesian scoring method using a constant prior with a mean of 0 and a standard deviation of 1.0. The three ICC response models were (1) the one-parameter model, in which test items differ only in terms of their difficulties (Rasch, 1960); (2) the two-parameter model, in which items may differ in terms of their difficulties and discriminations (Lord & Novick, 1968); and (3) the three-parameter model (Lord & Novick, 1968), in which items may differ in terms of difficulties, discriminations, and "guessing" parameters.

Test Data

Data used were from three different sources: (1) the hypothetical test and the structured set of response patterns used by Bejar and Weiss (1979), (2) a conventional classroom achievement test, and (3) a computer-administered adaptive achievement test.

Hypothetical response patterns. Using the example provided by Bejar and Weiss (1979), achievement level estimates were obtained for each possible response pattern to a hypothetical five-item test for which the parameters for each of the three response models were assumed to be known. The parameter values for the hypothetical test using the three-parameter model are shown in Table 1. All 32 possible response patterns were generated for these five items (see Table 2). Since M-L scoring methods cannot score response patterns with all items answered correctly or all items answered incorrectly, analyses were confined to the 30 response patterns scorable by all three scoring methods.

Table 1
Item Parameters for a Hypothetical Five-Item Test Assuming a Three-Parameter ICC Response Model

Item	Discrimination (a)	Difficulty (b)	Lower Asymptote (c)
1	1.00	-2.00	.10
2	1.50	-1.00	.10
3	1.00	0.00	.10
4	1.50	1.00	.10
5	1.00	2.00	.10

Conventional test. Data were obtained from the administration of a conventional classroom achievement test to a group of 200 undergraduate college students in an introductory biology course at the University of Minnesota. Estimates of the parameters of the three-parameter ICC model were available for 39 of the 55 items administered in this particular examination (see Bejar, Weiss, & Kingsbury, 1977).

The item parameter estimates were obtained using a method operationalized by Urry (1976). The procedure performs a direct conversion of the classical item parameters to obtain estimates of the discrimination (a) and difficulty (b) parameters and uses the value that minimizes a χ^2 statistic as an estimate of the "guessing" (c) parameter. Estimates are further refined by an ancillary correction procedure. Estimates of the parameter values for this examination were based on the responses of approximately 1200 people to each item. Final parameter estimates are shown in Appendix Table A.

Adaptive test. To determine whether the process of adapting a test to an individual's level of achievement might also affect the extent to which the different scoring methods yielded similar achievement level estimates for a group of individuals, additional data were obtained from the live administration of a computerized stratified adaptive (stradaptive) test. Utilizing the item pool from which the conventional test was drawn, this test was administered to a group of 200 volunteer students from the same biology course (Bejar, Weiss, & Gialluca, 1977).

The parameter estimates for the items in the stradaptive item pool were obtained from previous administrations of conventional classroom examinations. The ICC item parameter estimation procedure was the same as that used for the conventional test. The number of individuals on which the parameter estimates were based ranged from 638 to 998, depending on the original time of administration of the item. The parameters of the items in the stradaptive item pool are shown in Appendix Table B. The stradaptive test used a variable termination rule which terminated the test when an individual's ceiling stratum (Weiss, 1974, p. 46) had been identified. Test lengths actually taken by individuals varied from a minimum of 9 items to the maximum of 50 items.

Scoring and Analysis

Hypothetical test. Each of the 32 response patterns was scored by each of the three scoring methods (M-L normal, M-L logistic, and Bayesian) using the parameter values from Table 1. This represented an application of the three-parameter model. In order to use the two-parameter model, each of the response vectors was again scored with each scoring method; but the value of c for each item was set to zero (values of a and b for each item remained the same as in Table 1). To apply the one-parameter model, each response pattern was again scored by each scoring method; but the value of a for each item was set equal to 1.00, and the value of c was set to zero (values of b again remained as in Table 1).

To determine the extent to which the scoring method employed in achievement level estimation affected the rank ordering of the 32 response patterns, two analyses were performed. First, for each response model, differences among the scoring methods were examined by determining for each pair of scoring methods (1) the number of response patterns which were given different rankings, (2) the magnitude of the greatest difference in ranking, and (3) the average difference in ranking across all response patterns. Secondly, the degree of agreement among the scoring methods was quantified by obtaining values of Kendall's Tau (a rank order correlation coefficient) between achievement level estimates obtained from each pair of scoring methods within each response model. To the extent to which these correlations differed from

1.0, the scoring methods involved may be said to give divergent rankings of the same response patterns.

Conventional and adaptive tests. Conventional and adaptive test response patterns from the 200 subjects were scored by each of the three scoring methods at various points in the test. Scores were obtained after each three-item block in the test. Thus, this procedure produced scores based on the administration of 3 through 39 items in the conventional test and 3 through 48 items in the adaptive test, in increments of 3 items. This scoring was done first under the assumption of the three-parameter model, using the available item parameter estimates from Appendix Tables A and B. To investigate scoring by the two-parameter model, the scoring procedure described above was again employed (i.e., all response patterns were scored by each of the three scoring methods at each of a number of different test lengths). However, the parameters were edited so that although a and b for each item remained the same as in Appendix Tables A or B, c for each item was set to zero. Scoring by the one-parameter model was also done at 3-item increments for each test; but item parameter values were edited so that a for each item was set equal to 1.00, c for each item was set equal to zero, and b for each item remained as in Tables A or B.

Correlations were then calculated separately for the one-, two-, and three-parameter data between achievement level estimates generated by each pair of scoring methods at each of the 13 different test lengths between 3 and 39 items for the conventional test, and at each of the 16 different test lengths from 3 to 48 items for the adaptive test. To the extent that any correlation differed from 1.0, it might be said that at that particular test length the two scoring methods gave achievement level estimates that differed by more than a linear transformation.

Results

Hypothetical Test

One-parameter model. The achievement level estimates obtained for each of the possible response patterns from each of the scoring methods, assuming a one-parameter ICC response model, are shown in Table 2. The response patterns in which all items were answered correctly [1,1,1,1,1] and in which all items were answered incorrectly [0,0,0,0,0] have been omitted because the M-L estimates for these response patterns are positive and negative infinity, respectively. To make the comparison among scoring methods easier, the estimates have been ordered in terms of the ranking of the Bayesian achievement level estimates.

For the one-parameter model, the Bayesian achievement level estimates differed from the M-L normal estimates in rank order for 17 of the 30 response patterns. The average difference in ranking of a response pattern between the two methods was .43. The greatest difference in ranking between scores derived from the two models was a difference of 1.5 ranks.

The Bayesian estimates differed from the M-L logistic estimates in rank order for 28 of the 30 response patterns. The average difference in rank order was 2.07. The largest difference in ranking was 4.5 positions. This result

was confounded, however, by the large number of tied ranks obtained by the M-L logistic scoring method; there were only 4 unique scores for the 30 response patterns. By contrast, the Bayesian method gave unique θ estimates to all 30 response patterns.

Table 2
Achievement Level Estimates and Rank Orders for Bayesian and Maximum-Likelihood (M-L) Scoring Methods Assuming a One-Parameter ICC Response Model

Response Pattern[a]	Bayesian		M-L Normal		M-L Logistic	
	Estimate	Rank	Estimate	Rank	Estimate	Rank
1,1,1,0,1	1.05	1	1.59	2	1.61	3
1,1,0,1,1	1.00	2	1.38	3	1.61	3
1,1,1,1,0	.97	3	1.62	1	1.61	3
1,0,1,1,1	.84	4	1.09	4	1.61	3
0,1,1,1,1	.58	5	.79	5	1.61	3
1,1,0,0,1	.54	6	.69	6	.51	10.5[b]
1,0,0,1,1	.47	7	.51	8.5[b]	.51	10.5
1,0,1,0,1	.42	8	.51	8.5	.51	10.5
1,1,0,1,0	.38	9	.51	8.5	.51	10.5
1,1,1,0,0	.34	10	.51	8.5	.51	10.5
0,1,0,1,1	.27	11	.29	12	.51	10.5
1,0,1,1,0	.27	12	.33	11	.51	10.5
0,0,1,1,1	.24	13	.20	14	.51	10.5
0,1,1,0,1	.21	14	.26	13	.51	10.5
0,1,1,1,0	.07	15	.07	15	.51	10.5
1,0,0,0,1	.01	16	-.07	16	-.51	20.5
0,0,0,1,1	-.03	17	-.20	17	-.51	20.5
0,1,0,0,1	-.15	18	-.26	18	-.51	20.5
0,0,1,0,1	-.16	19	-.29	19	-.51	20.5
1,0,0,1,0	-.21	20	-.33	20	-.51	20.5
1,0,1,0,0	-.33	21	-.51	22.5	-.51	20.5
1,1,0,0,0	-.33	22	-.51	22.5	-.51	20.5
0,0,1,1,0	-.34	23	-.51	22.5	-.51	20.5
0,1,0,1,0	-.35	24	-.51	22.5	-.51	20.5
0,1,1,0,0	-.47	25	-.69	25	-.51	20.5
0,0,0,0,1	-.53	26	-.79	26	-1.61	28
0,0,0,1,0	-.78	27	-1.09	27	-1.61	28
0,0,1,0,0	-.97	28	-1.38	28	-1.61	28
1,0,0,0,0	-1.02	29	-1.62	30	-1.61	28
0,1,0,0,0	-1.05	30	-1.59	29	-1.61	28

[a]The response patterns [0,0,0,0,0] and [1,1,1,1,1] are not included because M-L estimates cannot be obtained for these response patterns.

[b]Ties were assigned the average of the ranks that the tied estimates would span if they were not tied.

The ranks of the M-L normal estimates differed from those of the M-L logistic method for 28 of the 30 response patterns. The average difference

in rank order was 2.00, and the maximum difference in ranking was 4.5. Again, the small number of unique ranks assigned by the M-L logistic method partially accounted for this difference; the M-L normal method gave unique θ estimates to 24 of the 30 response patterns.

It is evident from these data that using the one-parameter model, the three scoring methods resulted in different θ estimates. Although there were only relatively small differences in the rank ordering of the θ estimates between the Bayesian and the M-L normal methods, all θ estimates generated by the Bayesian method were uniformly closer to zero than those of the M-L normal method. The differences were particularly large at the extremes, where the differences were as much as .50 score units on the achievement metric for the [1,1,1,0,1] and [0,1,0,0,0] response patterns. The tendency of the Bayesian θ estimates to be closer to zero was also evident in comparison to the M-L logistic method. However, because of the tendency of the M-L logistic method not to provide different θ estimates for different response patterns, differences approaching .50 units were evident between the two methods for response patterns obtaining θ estimates near the mean (e.g., response pattern [1,0,0,0,1]).

Using the one-parameter model, the M-L logistic scoring method resulted in different θ estimates for different numbers of items answered correctly. Thus, θ estimates of 1.61 were obtained for all response patterns in which only 4 items were answered correctly; θ estimates of .51 were given to all response patterns in which 3 items were answered correctly; θ estimates of -.51 were obtained for all patterns with 2 correct answers; and θ estimates of -1.61 were assigned to all patterns with only 1 correct answer. It should be noted that the items were all of differing difficulties (see Table 1). Thus, the one-parameter M-L logistic scoring method provides θ estimates based on the number of items answered correctly, but does not take into account the difficulties of the items; all response patterns with the same number-correct score will result in the same θ estimates, regardless of whether easy or difficult items are answered correctly. This property of the one-parameter M-L logistic scoring method is the basis for the use of number-correct score in the Rasch (1960) one-parameter logistic ICC model. By contrast, both the M-L normal and Bayesian scoring methods resulted in different θ estimates for items of differing difficulty; in these scoring methods the difficulties of items answered correctly or incorrectly are taken into account in estimating θ levels.

Two-parameter model. The estimates of achievement level for all the possible response patterns (except [0,0,0,0,0] and [1,1,1,1,1]) for the two-parameter response model are shown in Table 3; for these data the Bayesian estimates differed from the M-L normal estimates in terms of rank order in 16 of 30 instances. The average difference in rank position between the two methods was .65; the maximum difference in the ranking of the two methods was a difference of 3 positions.

The Bayesian estimates differed from the M-L logistic estimates in rank order for 28 of the 30 response patterns, and the average difference in rank position was 1.93. The maximum difference in rank was 4.5 positions.

The M-L normal estimates differed from the M-L logistic estimates in terms of rank order for 28 of the 30 response patterns, and the average

difference in rank position was 1.63. The largest discrepancy in the rankings was a difference of 4.5 positions.

Table 3
Achievement Level Estimates and Rank Orders for Bayesian and Maximum-Likelihood (M-L) Scoring Methods Assuming a Two-Parameter ICC Response Model

Response Pattern[a]	Bayesian		M-L Normal		M-L Logistic	
	Estimate	Rank	Estimate	Rank	Estimate	Rank
1,1,0,1,1	1.09	1	1.42	2	1.60	2
1,1,1,1,0	1.08	2	1.63	1	1.60	2
1,1,1,0,1	.93	3	1.24	3	1.19	4.5
0,1,1,1,1	.64	4	.93	4	1.60	2
1,1,0,1,0	.63	5	.78	5	.84	7
1,0,1,1,1	.62	6	.61	6	1.19	4.5
1,1,0,0,1	.51	7	.60	7	.46	11.5
0,1,0,1,1	.41	8	.50	8	.84	7
1,0,0,1,1	.39	9	.30	11	.46	11.5
1,1,1,0,0	.31	10	.42	9	.46	11.5
0,0,1,1,1	.30	11	.13	14	.46	11.5
0,1,1,1,0	.28	12	.39	10	.84	7
1,0,1,1,0	.23	13	.17	13	.46	11.5
0,1,1,0,1	.17	14	.23	12	.46	11.5
1,0,1,0,1	.11	15.5[b]	.03	15	.00	15.5
0,0,0,1,1	.11	15.5	-.13	17	-.46	19.5
0,1,0,1,0	.00	17	-.03	16	.00	15.5
1,0,0,1,0	-.06	18	-.17	18	-.46	19.5
0,0,1,1,0	-.11	19	-.30	20	-.46	19.5
0,1,0,0,1	-.15	20	-.23	19	-.46	19.5
1,0,0,0,1	-.24	21	-.39	21	-.84	24
0,0,1,0,1	-.28	22	-.50	23	-.84	24
1,1,0,0,0	-.29	23	-.42	22	-.46	19.5
0,0,0,1,0	-.38	24	-.61	25	-1.19	26.5
0,1,1,0,0	-.42	25	-.60	24	-.46	19.5
1,0,1,0,0	-.58	26	-.78	26	-.84	24
0,0,0,0,1	-.64	27	-.93	27	-1.60	29
0,1,0,0,0	-.89	28	-1.24	28	-1.19	26.5
0,0,1,0,0	-1.06	29	-1.42	29	-1.60	29
1,0,0,0,0	-1.16	30	-1.63	30	-1.60	29

[a]The response patterns [0,0,0,0,0] and [1,1,1,1,1] are not included because M-L estimates cannot be obtained for these response patterns.

[b]Ties were assigned the average of the ranks that the tied estimates would span if they were not tied.

As in the case of the one-parameter model, it was again apparent that the three scoring methods resulted in different estimates of achievement levels. Estimates obtained from the Bayesian method showed the same tendency toward more moderate estimates (i.e., estimates closer to zero) that was exhibited

using the one-parameter model. This result occurred when the Bayesian scoring method was compared with either of the M-L scoring methods. The magnitude of the discrepancies between the Bayesian estimates and the M-L normal estimates was almost exactly the same as with the one-parameter model. Comparison between the Bayesian estimates and the M-L logistic estimates was again made difficult by the fact that the M-L logistic method sorted the 30 response patterns into only 9 different achievement levels. However, differences between the estimates appeared to be greater for response patterns which received extreme achievement estimates than for those which received moderate estimates.

The observation that the M-L logistic method yielded 9 different achievement levels indicates that the number of correct responses is no longer a sufficient description of the M-L logistic achievement level estimate using the two-parameter model. In fact, as the data in Table 3 indicate, the sufficient indicant of the M-L logistic achievement level estimate using the two-parameter model was the discrimination of the items answered incorrectly in a testee's response pattern. This finding has been reported earlier by Samejima (1969) and indicates that the difficulty of items answered correctly or incorrectly has no effect on achievement level estimates obtained using the two-parameter M-L logistic scoring method.

Three-parameter model. The estimates of achievement level for each of the response patterns when a three-parameter item characteristic response model was assumed are shown in Table 4. It may be seen from this table that the M-L normal scoring algorithm failed to converge on an estimate for 7 of the 30 response patterns. The M-L logistic algorithm failed for 9 of the 30 patterns. These failures occurred when the likelihood function was too flat to allow the algorithm (a Newton-Raphson procedure; see Bejar & Weiss, 1979, pp. 10-11) to determine the point of maximization within 100 attempts. In this test the likelihood function was flattened because of the addition of the lower asymptote parameter, c, the "pseudo-guessing" parameter. The effect of this parameter is to lower the amount of information obtained from any single response, thereby flattening the likelihood function.

For both M-L scoring methods the nonconvergences occurred for the 6 response patterns which were given the lowest θ estimates by the Bayesian method (the value of -8.77 for the M-L normal method represents an artificial convergence). In addition, both M-L methods failed for the [0,1,0,1,1] response pattern, which represents the responses of an individual who answered easy items (Items 1 and 3) incorrectly and difficult items (Items 4 and 5) correctly. The M-L logistic scoring method also failed to converge for the [0,1,0,1,0] response pattern, in which incorrect responses were given to the items with lower discriminations and correct responses were given to the higher discriminating items. As Table 4 shows, because the Bayesian scoring method does not use an iterative procedure, θ estimates were obtained for all 30 response patterns.

Due to these convergence failures, it was appropriate to examine the differences in the three scoring methods' rankings by including in the rankings only those response patterns for which θ estimates were obtained by all three methods. These curtailed rankings are shown as Rank 2 in Table 4.

Table 4
Achievement Level Estimates and Rank Orders for Bayesian and Maximum-Likelihood (M-L) Scoring Methods Assuming a Three-Parameter ICC Response Model

Response Pattern[a]	Bayesian			M-L Normal			M-L Logistic		
	Estimate	Rank	Rank 2[b]	Estimate	Rank	Rank 2	Estimate	Rank	Rank 2
1,1,1,1,0	.91	1	1	1.58	1	1	1.56	1	1
1,1,0,1,1	.60	2	2	1.20	2	2	1.34	2	2
1,1,1,0,1	.53	3	3	.98	3	3	.89	4	4
1,1,1,0,0	.23	4	4	.37	5	5	.41	7	7
1,1,0,1,0	.16	5	5	.58	4	4	.58	5	5
0,1,1,1,1	.02	6	6	-.59	8	8	1.33	3	3
1,1,0,0,1	-.15	7	7	-.33	6	6	-.35	8	8
0,1,1,1,0	-.27	8	8	-.71	9	9	.51	6	6
1,1,0,0,0	-.33	9	9	-.47	7	7	-.49	9	9
1,0,1,1,1	-.33	10	10	-.96	12	12	-.99	12	12
0,1,1,0,1	-.49	11	11	-.77	10	10	-.57	10	10
1,0,1,1,0	-.53	12	12	-.99	13	13	-1.06	13	13
1,0,1,0,1	-.69	13	13	-1.01	14	14	-1.09	14	14
0,1,1,0,0	-.60	14	14	-.82	11	11	-.79	11	11
1,0,1,0,0	-.77	15	15	-1.03	15	15	-1.14	15	15
0,1,0,1,1	-.83	16	--	NC[c]	--	--	NC	--	--
0,1,0,1,0	-.92	17	--	-2.31	22	--	NC	--	--
0,1,0,0,1	-1.00	18	16	-1.45	16	16	-1.44	16	16
1,0,0,1,1	-1.04	19	17	-1.68	19[d]	19	-1.60	18	18
0,1,0,0,0	-1.05	20	18	-1.46	17	17	-1.50	17	17
1,0,0,1,0	-1.09	21	19	-1.68	19	19	-1.63	19.5	19.5
1,0,0,0,1	-1.15	22	20	-1.68	19	19	-1.63	19.5	19.5
1,0,0,0,0	-1.17	23	21	-1.69	21	21	-1.65	21	21
0,0,1,1,1	-1.31	24	--	NC	--	--	NC	--	--
0,0,1,1,0	-1.35	25	--	NC	--	--	NC	--	--
0,0,1,0,1	-1.39	26	--	NC	--	--	NC	--	--
0,0,1,0,0	-1.42	27	--	-8.77	23	--	NC	--	--
0,0,0,1,1	-1.70	28	--	NC	--	--	NC	--	--
0,0,0,1,0	-1.71	29	--	NC	--	--	NC	--	--
0,0,0,0,1	-1.72	30	--	NC	--	--	NC	--	--

[a]The response patterns [0,0,0,0,0] and [1,1,1,1,1] are not included because M-L estimates cannot be obtained for these response patterns.

[b]Ranking of response patterns for which all three methods obtained estimates.

[c]The M-L estimation algorithm failed to converge on a unique maximum.

[d]Ties were assigned the average of the ranks that the tied estimates would span if they were not tied.

Using these curtailed rankings, the Bayesian estimates differed in rank order from the M-L normal estimates for 15 of 21 response patterns. The average difference in rank position between the two methods was .95. The largest difference in ranks was 3. The Bayesian estimates also differed from the M-L logistic estimates for 14 of 21 response patterns. The average difference in ranks between these methods was .95 ranks, and the maximum difference was 3.

The M-L normal ranking differed from the M-L logistic ranking for 10 of 21 response patterns. The average difference between the rankings of the estimates derived from the two scoring method rankings was .81. The largest difference in rank order was 5.

The most obvious effect of the addition of the third parameter was that the achievement level estimates obtained by each of the three scoring methods were consistently lower than those obtained using the one- and two-parameter models. This result may be explained by the fact that the third parameter indicates the ease with which an item might be answered correctly without any knowledge of the subject matter. As the level of this parameter increases, the weight given to a correct answer is decreased for each of the scoring methods; therefore, the final θ estimates are lower.

For the response patterns for which each of the scoring methods obtained an achievement level estimate, the tendency for the Bayesian scoring method to result in more moderate estimates than either of the M-L methods was still evident, as it was under the one- and two-parameter models. Also, the tendency for the discrepancies between the estimates to be higher for response patterns in which the estimates were quite different from zero was still apparent, particularly in the comparison between the Bayesian method and the M-L normal method. For example, for the 3 response patterns giving rise to the most extreme θ estimates--[1,0,0,1,0], [1,0,0,0,1], and [1,0,0,0,0]--the average difference between the estimates was .55 score units; for the 3 response patterns for which the θ estimates were closest to zero--[1,1,0,1,0], [0,1,1,1,1], and [1,1,0,0,1]--the average difference between the estimates was .41 score units.

The M-L logistic estimates using the three-parameter model were not as obviously related to the discriminations of items answered incorrectly as in the two-parameter data. Thus, the three-parameter data permitted the first clear comparison of the differences between the Bayesian and M-L logistic estimates. In general, the Bayesian θ estimates tended to be less extreme (e.g., closer to zero) than the M-L logistic θ estimates, similar to the comparison between the Bayesian and M-L normal estimates. However, there was no trend for the estimates for the response patterns with extreme θ estimates to diverge to a greater extent than those with moderate θ estimates, as in the comparison between the Bayesian and M-L normal estimates.

Relationships among models and methods. Values of Kendall's Tau among achievement level estimates generated by the three scoring methods within each response model are shown in Table 5. The highest correlation between scoring methods was between the Bayesian method and the M-L normal method for both the one-parameter and two-parameter models (Tau=.963 and .948, respectively). For the three-parameter model, the most similar ranks were obtained by the two M-L methods (Tau=.918). For all three models, the least similar sets of rankings were derived from the Bayesian and M-L logistic methods. When the second and third parameters were added to the response models, there was a tendency for the correlations between pairs of scoring methods to become more similar as the correlations between the M-L logistic ranks and those of the other two scoring methods increased. At the same time, there was a decrease in the similarity of rankings produced by the Bayesian and M-L normal methods. Using the three-parameter model, the three pairs of correla-

tions tended to cluster around a Tau of .90, accounting for about 81% common variance in the pairs of rankings produced by the three scoring methods.

Table 5
Values of Kendall's Tau Among Achievement Estimates from Three Scoring Methods for Each ICC Response Model

Scoring Methods	Response Model		
	One-Parameter	Two-Parameter	Three-Parameter
Bayesian vs. M-L Normal	.963	.948	.906
Bayesian vs. M-L Logistic	.864	.873	.893
M-L Normal vs. M-L Logistic	.876	.898	.918

Conventional Test

Convergence failures. The data from the hypothetical test indicated that the M-L scoring methods failed to obtain achievement level estimates under certain circumstances. M-L scoring methods will be unable to converge for response patterns which include either all correct answers or all incorrect answers. In addition, there were other response patterns with likelihood functions that did not have a single obvious maximum. These kinds of response patterns will also result in convergence failures.

Table 6
Percentage of Maximum-Likelihood Convergence Failures for Conventional Test Data with Varying Numbers of Items (N=200)

Number of Items	Percentage of Convergence Failures					
	One-parameter model		Two-parameter model		Three-parameter model	
	M-L Normal	M-L Logistic	M-L Normal	M-L Logistic	M-L Normal	M-L Logistic
3	63	63	63	63	66	65
6	27	27	27	27	29	30
9	17	17	17	17	17	17
12	13	13	13	13	13	13
15	10	10	10	10	10	10
18	8	8	8	8	8	8
21	8	8	8	8	8	8
24	6	6	6	6	6	6
27	5	5	5	5	5	5
30	4	4	4	4	4	4
33	4	4	4	4	4	4
36	1	1	1	1	1	1
39	1	1	1	1	1	1

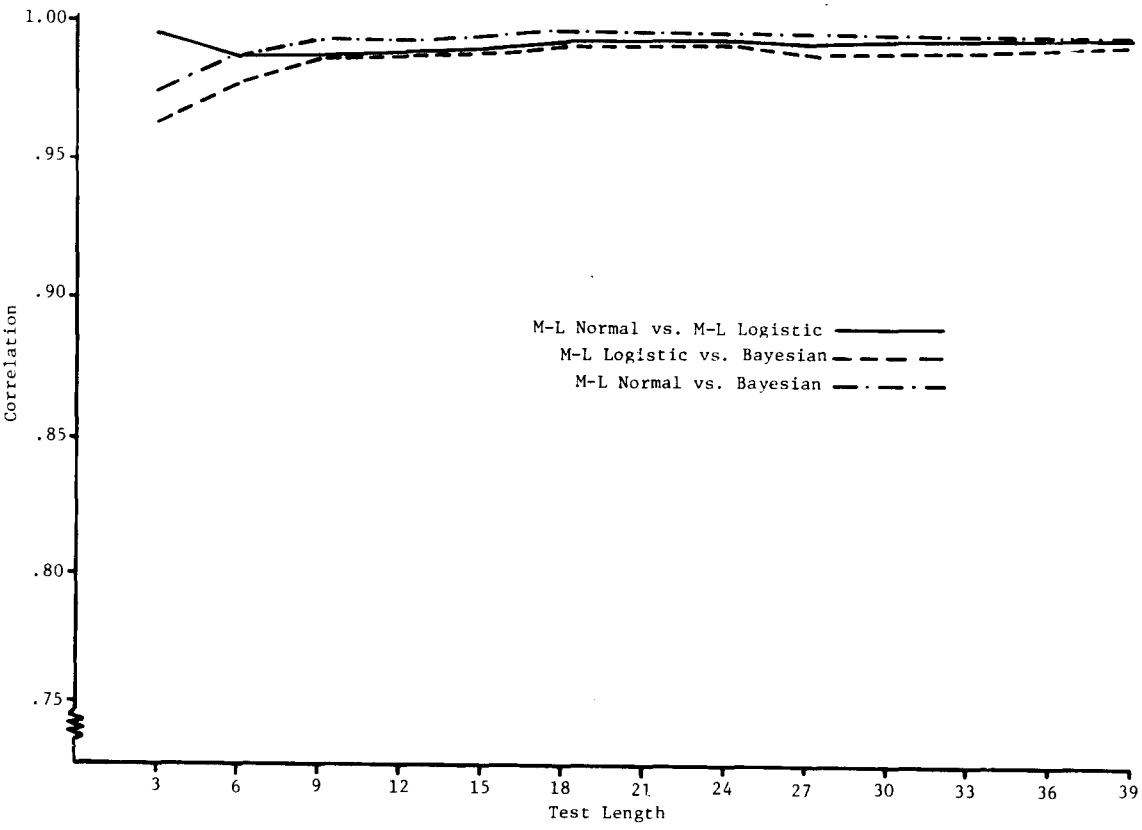
Table 6 shows the percentage of individuals for whom the M-L scoring methods did not converge on a unique achievement level estimate for each test

length and response model, using conventional test response data. The M-L scoring methods failed to obtain achievement level estimates for almost two-thirds of the response patterns at the shortest test length (3 items), regardless of the response model or the scoring method used. At a test length of 6 items, the convergence failure rate varied between 27% and 30% of the response patterns. For both 3-item and 6-item tests, there were no differences in the percentage of convergence failures between the M-L normal and M-L logistic scoring methods within the one-parameter and two-parameter models. Similarly, there were no differences between these two models regardless of scoring method. For both M-L logistic and M-L normal scoring methods, the three-parameter model resulted in slightly more convergence failures than the one- and two-parameter models, for 3- and 6-item tests.

For conventional tests of 9 or more items, there were no differences among models or methods of scoring in the rate of convergence failures. The percentage of convergence failures dropped consistently with increasing test length. But even for relatively long tests (e.g., 30 items), 4% of the 200 response patterns failed to converge within 100 iterations. At the longest test length (39 items), 1% of the response patterns failed to yield convergent estimates for all methods and models of M-L scoring.

One-parameter model. Appendix Table C shows Pearson product-moment correlations among scores derived from each pair of the three scoring methods for test lengths of 3 to 39 items, in steps of 3 items; these correlations were

Figure 1
Correlations Between Achievement Level Estimates as a Function of Test Length for Conventional Test Data Using a Two-Parameter Model

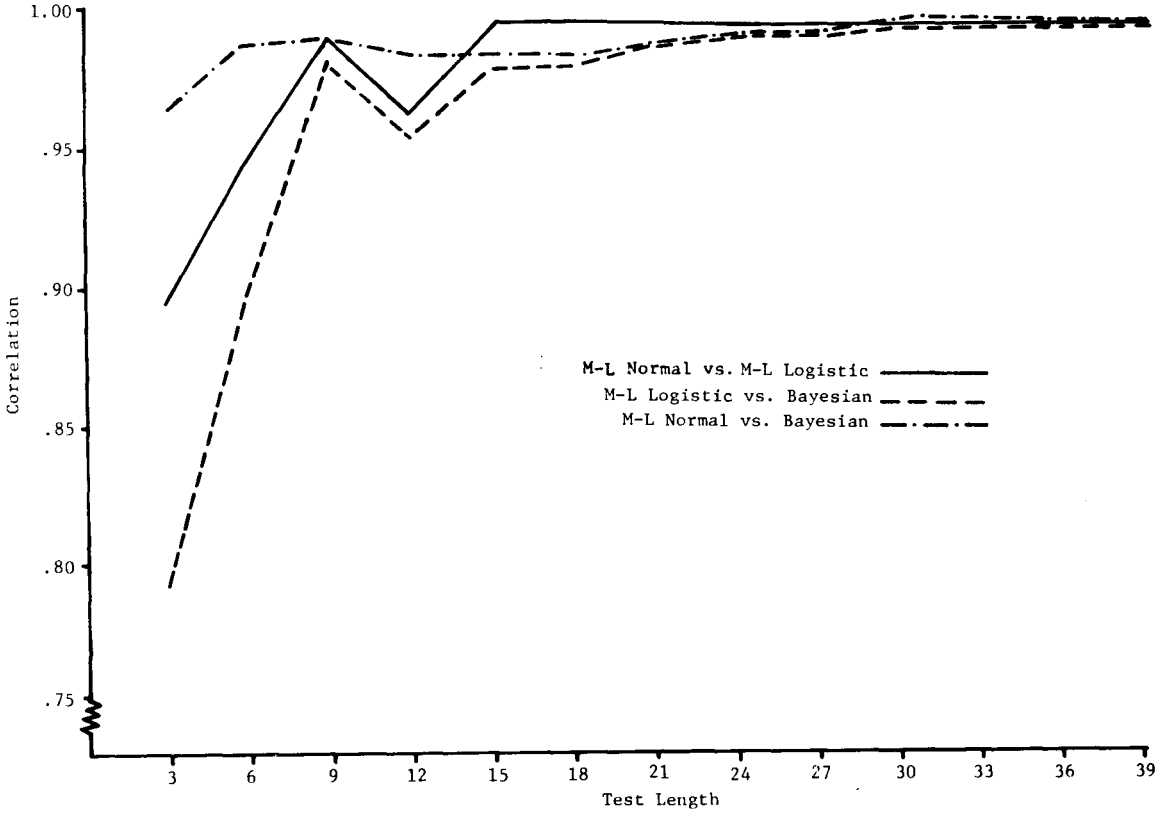


based on only those cases for which the M-L scoring estimates converged. As the data show, the minimum correlation was r=.9741 for scores from the M-L logistic and Bayesian methods for a 3-item test. The maximum r was .9967 for scores from the M-L normal and Bayesian methods for an 18-item test. There was no general trend in the data either as a function of test length or scoring method. In all cases, for tests greater than 3 items, more than 97% of the variance in a scoring method was common with the other scoring methods.

Two-parameter model. Figure 1 shows the correlations between scores derived from the three scoring methods when the data were scored by the two-parameter model (numerical values are in Appendix Table C). In general, the correlations were slightly lower than when the data were scored using only the difficulty parameter information. For the two-parameter data, the minimum correlation was .9629 between the M-L logistic and Bayesian methods, at a test length of 3 items. The highest correlation was .9958 between the M-L normal and M-L logistic methods for a 3-item test. As Figure 1 shows, there was a slight trend toward higher correlations as test length increased. For the two-parameter data, 97% of the variance in scores was common between all pairs of methods for test lengths greater than 6 items.

Three-parameter model. Figure 2 shows the correlations among the achievement level estimates obtained from each of the scoring methods at test lengths from 3 to 39 items when the data were scored using a three-parameter ICC response model (numerical values are in Appendix Table C). It can be seen

Figure 2
Correlations Between Achievement Level Estimates as a Function of Test Length for Conventional Test Data Using a Three-Parameter Model



from Figure 2 that the correlations among the three scoring methods were considerably lower for the three-parameter model at test lengths of 15 items or less than they were when only one or two parameters were used to score the data. The lowest correlation was r=.7917 for the M-L logistic versus Bayesian comparison for tests of 3 items; the highest correlation was r=.9967 for the M-L normal versus M-L logistic comparison for tests of 39 items. The lowest correlations occurred uniformly for 3-item tests, with large increases into the r=.90 range for all correlations for 6-item tests. There was a general trend for all correlations to increase with increasing test length, except for a slight drop at 12 items associated with the M-L logistic method. There were only very small differences among correlations at test lengths of 27 or more items. There was a general tendency throughout the data for scores from the M-L logistic and Bayesian methods to correlate lowest, with the trend most pronounced at shorter test lengths. For the three-parameter data, 97% of the variance in each scoring method was common with the other scoring methods for tests 15 items or more in length.

Summary. The data show a general decrease in similarity among scores as more parameters were used to score the items. The addition of the discrimination parameter tended to reduce correlations among scoring methods slightly for tests of less than 9 items in length; however, there were no large differences between scoring methods for the two-parameter data. When the "guessing" parameter was added, there was a marked decrease in similarity among scores associated with the M-L logistic method for tests shorter than 18 items; relationships between the M-L normal scores and the Bayesian scores remained high, although they were somewhat lower for most test lengths than with two-parameter scoring.

Adaptive Test

Convergence failures. Table 7 shows the percentage of response patterns for which the M-L scoring methods failed to obtain an achievement level estimate at each test length from 3 to 48 items using each response model. These data show that there were no consistent differences between the M-L logistic and M-L normal scoring methods and no differences at all between these methods using the one- and two-parameter response models.

Under each response model, 20 to 38% of the response patterns resulted in estimation failures for the shortest test length. Fewer estimation failures were noted at longer test lengths. For the one- and two-parameter models, no convergence failures were observed for any test length greater than 9 items. Under the assumption of the three-parameter model, more convergence failures were noted than for the simpler response models for test lengths up to 33 items. No convergence failures were observed at any test length greater than 33 items.

These results were not completely comparable to convergence failures observed for the conventional test because of the stradaptive variable length termination. At longer test lengths the number of testees on which the percentages were based dropped steadily as the ceiling stratum for individuals was determined. This variable termination criterion may add an unknown amount of bias to comparisons made between the conventional and adaptive tests in this study.

Table 7
Percentage of Maximum Likelihood Convergence Failures for Adaptive Test Data with Varying Numbers of Items

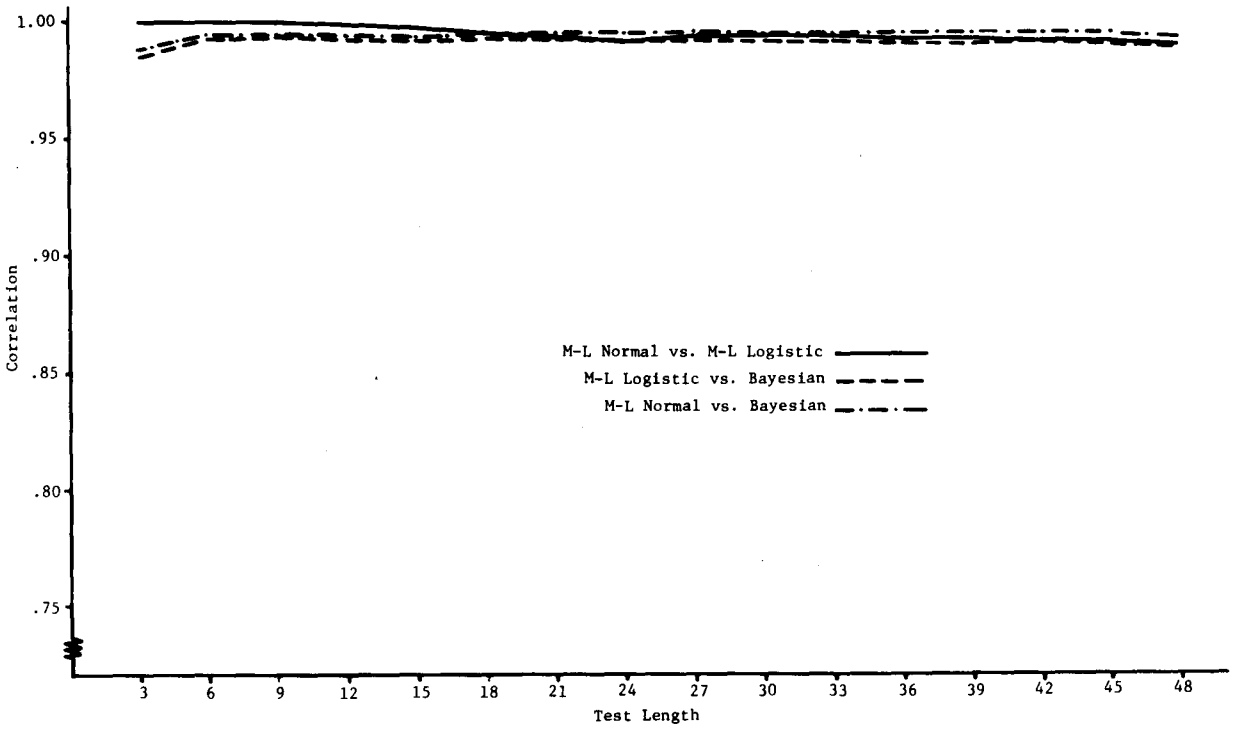
Number of Items	Number of Individuals	Percentage of Convergence Failures: One-parameter model M-L Normal	One-parameter model M-L Logistic	Two-parameter model M-L Normal	Two-parameter model M-L Logistic	Three-parameter model M-L Normal	Three-parameter model M-L Logistic
3	200	20	20	20	20	38	30
6	200	6	6	6	6	9	11
9	200	1	1	1	1	4	6
12	185	0	0	0	0	1	2
15	169	0	0	0	0	1	2
18	143	0	0	0	0	1	2
21	127	0	0	0	0	2	3
24	108	0	0	0	0	0	0
27	97	0	0	0	0	1	1
30	83	0	0	0	0	2	1
33	79	0	0	0	0	1	0
36	67	0	0	0	0	0	0
39	60	0	0	0	0	0	0
42	56	0	0	0	0	0	0
45	51	0	0	0	0	0	0
48	47	0	0	0	0	0	0

One-parameter model. Appendix Table D shows Pearson product-moment correlations between achievement level estimates derived from each pair of the three scoring methods for test lengths of 3 to 48 items. These correlations were based only on those individuals for whom the M-L scoring methods did not fail to converge and for whom the test continued to the specified test length. The data show that the lowest observed correlation was .9927 for scores from the M-L logistic and Bayesian methods for a test length of 3 items. The highest observed correlation was .9998, between scores from the M-L logistic and M-L normal methods at the 9-item test length and from the M-L normal and Bayesian methods at all test lengths between 24 and 45 items. For all test lengths, more than 97% of the score variance for each scoring method was common with every other scoring method.

Two-parameter model. Figure 3 shows the correlations between achievement level estimates derived from each pair of the three scoring methods as a function of test length, assuming a two-parameter response model (numerical values are shown in Appendix Table D). These correlations were, in general, slightly lower than those observed under the one-parameter model. The lowest observed correlation was .9854, between scores obtained from the M-L logistic and Bayesian methods for a test length of 3 items. The highest observed correlation was .9996, between scores from the M-L logistic and M-L normal methods, also at a test length of 3 items. Again, at all test lengths, more than 97% of the score variance in a scoring method was common with every other method. As with the one-parameter model, no general trend was noted in the data as a function of test length, other than a very slight tendency for the

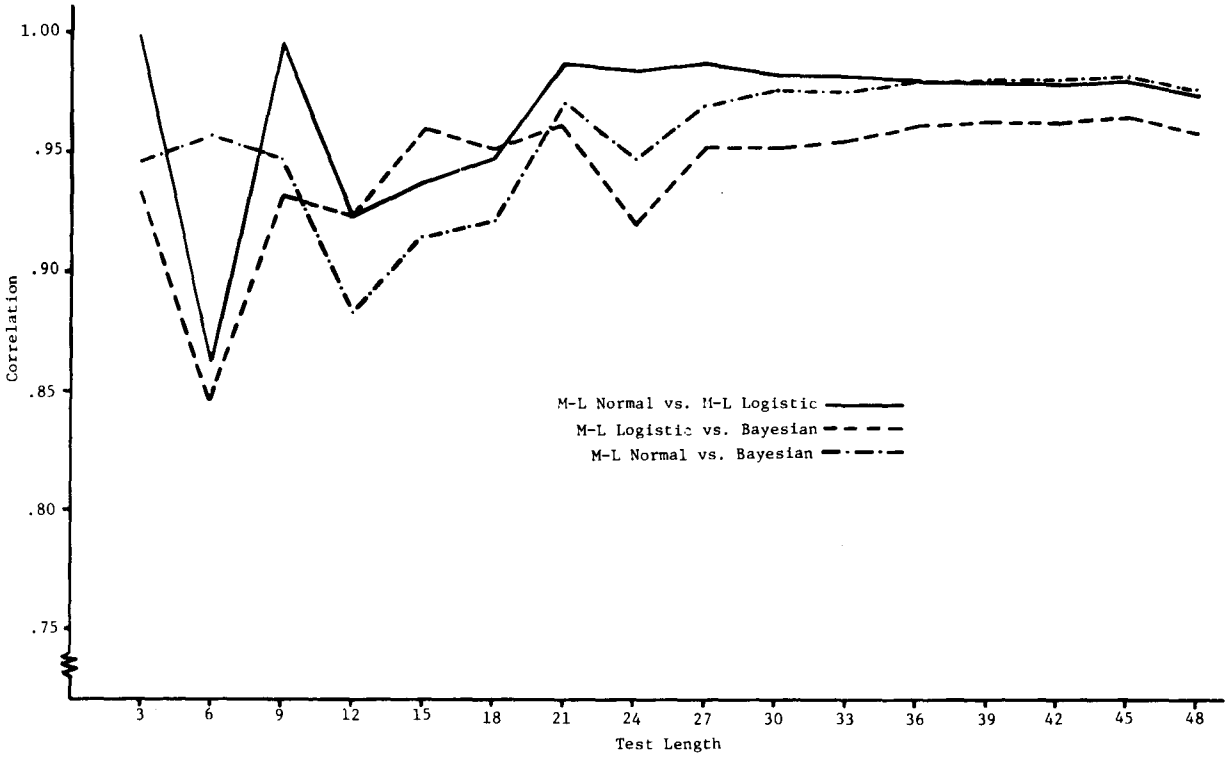
correlation between scores from the M-L normal and M-L logistic methods to decrease as the test length increased; but even at the longest test length observed (48 items), this correlation was still .9892. Figure 3 shows a slight tendency toward lower correlations between the Bayesian and M-L methods for the 3-item test length, followed by very consistent correlations at all longer test lengths.

Figure 3
Correlations Between Achievement Level Estimates as a Function of Test Length for Adaptive Test Data Using a Two-Parameter Response Model



Three-parameter model. Figure 4 shows the correlations between scores obtained from each pair of the three scoring methods as a function of test length for the three-parameter model (numerical values are in Appendix Table D). It is evident from this figure that the very consistent and high correlations observed under the assumption of the one- and two-parameter models were not observed when the three-parameter model was assumed, particularly for shorter test lengths. The lowest correlation observed under the assumption of the three-parameter model was .8444, between scores from the M-L logistic and Bayesian models at the 6-item test length. The highest correlation observed was .9997, between estimates from the M-L logistic and M-L normal methods at the 3-item test length. There was a general tendency for the correlations among the scores obtained from each pair of the three scoring methods to become higher and more consistent at longer test lengths. There was, however, no test length for which more than 97% of the score variance was common among the three scoring methods. This is the only combination of testing method and response model examined in this study for which this common variance criterion was not met at any test length.

Figure 4
Correlations Between Achievement Level Estimates as a Function of Test Length for Adaptive Test Data Using a Three-Parameter Model



At test lengths of 21 items or more, the M-L logistic and Bayesian scoring methods produced the least similar scores. For test lengths between 12 and 18 items, the lowest correlations were associated with the M-L normal and Bayesian scoring methods. Between 3 and 9 items, however, the lowest correlations were again associated with the M-L logistic and Bayesian comparison. Thus, these data show a general tendency for the Bayesian θ estimates to be consistently less similar to the M-L estimates than were the θ estimates for the two M-L scoring methods.

Summary. These data show a tendency toward greater dissimilarity among scores obtained from the three scoring methods when more complex response models were used to score the item responses from the adaptive test data. The use of a varying discrimination parameter in the two-parameter model reduced all observed correlations slightly (.0062 on the average), and the correlations between M-L logistic scores and Bayesian scores most noticeably (.0073 on the average). When a nonzero "guessing" parameter was used in the three-parameter model to obtain achievement level estimates, correlations among scores from the three different scoring methods decreased to a much greater extent (.0350 mean decrease), with the greatest decrease again being observed in correlations between scores from the M-L logistic and Bayesian methods (.0460 mean decrease). The three-parameter results showed less similarity among the scores obtained from the three scoring methods than either the one- or two-parameter results for each test length; differences among the achievement level estimates for

the one- and two-parameter models might be called unimportant, since correlations between the estimates were consistent for tests of reasonable lengths and tended to differ very little from 1.0. The three-parameter response model yielded consistently lower correlations between scores obtained using the three scoring methods; these correlations did not approach 1.0, even for long test lengths.

Comparison of Conventional and Adaptive Data

For the one-parameter model, correlations between scores obtained through the three different scoring methods were uniformly high; but those obtained from the adaptive testing procedure tended to be slightly higher than those obtained from the conventional testing procedure, for all test lengths. Using the one-parameter model with conventional test data, the average correlation observed between scores obtained from all pairs of scoring methods across all test lengths was .9920; for the adaptive test data, the average correlation was .9990.

Under the assumption of the two-parameter model, there was still a trend for the correlations between scores to be higher for data from the adaptive testing procedure than for data from the conventional testing procedure; but this trend was not as strong as that observed under the assumption of the one-parameter model. For the two-parameter model, the average observed correlation between scores from the three scoring methods across all test lengths for the conventional test was .9900. For the adaptive test data, the average correlation was .9929.

Under the assumption of the three-parameter model, the mean correlation between scores from the three scoring procedures for all test lengths was .9799 using responses to the conventional test and .9582 using responses to the adaptive test. Under this response model, the trend was for the scores obtained from the conventional test to be more consistent across the three scoring models than the scores obtained from the adaptive test. This trend is the opposite of the trend observed for the one- and two-parameter models.

One further point is of interest for the comparison of the adaptive and conventional testing procedures. Tables 6 and 7 show that the adaptive test data resulted in fewer M-L convergence failures than the conventional test data at every comparable test length. This difference resulted in 40% to 100% fewer observed estimation failures for the adaptive testing procedure. For the one- and two-parameter models, no estimation failures were observed at any test length greater than 9 items for the adaptive test data; for the conventional test data, estimation failures were observed at every test length up to 39 items, the longest test length examined. Using the three-parameter model, no estimation failures were observed at any test length greater than 33 items for the adaptive test data; but failures were observed for the conventional data up to the longest test length of 39 items.

Discussion and Conclusions

The data show that under certain conditions, the three ICC-based scoring methods will result in different achievement level estimates. Trends evident in the hypothetical test data were, in some cases, clarified by the analysis

of the conventional and adaptive test data. The data from the hypothetical five-item test clearly illustrated that θ estimates from the one-parameter logistic model scored by maximum likelihood are directly related to the number of items answered correctly, regardless of the difficulties of the items answered correctly or incorrectly. It is this property of the one-parameter logistic model which permits the Rasch model to use the number-correct score within an ICC framework. When all three scoring methods were applied to the same data, however, the results indicated that the M-L logistic scoring method in the one-parameter case ignored information that allowed differentiation among dissimilar response patterns having the same number-correct score. From an ICC point of view, promising fuller use of test response information, the one-parameter M-L logistic scoring method is no more informative than the number-correct score which it reflects, at least for short tests similar to the five-item hypothetical test. When the three scoring models were applied to live-testing data from both conventional and adaptive tests, correlations among θ estimates derived from the one-parameter model were quite high, regardless of test length. Thus, in the live-testing data, the fact that the M-L logistic scoring method ignored the item difficulties did not seriously affect its performance in comparison to the other two scoring methods.

When the hypothetical test data were scored using both the difficulty and discrimination parameters, the M-L logistic method still did not use the item difficulties in arriving at θ estimates. In this case, the M-L logistic θ estimates were associated, not with number-correct scores, but with the item discriminations; individuals who incorrectly answered items of the same discrimination, but with differing difficulties, all received the same θ estimate. Again, both the Bayesian and M-L normal scoring methods provided differential and highly correlated θ estimates, which took into account both the response pattern data and the item difficulties and discriminations. In live-testing data, in which all possible response patterns are unlikely to occur (as they did in the hypothetical test data), this trend again seemed to lack practical importance. In both the adaptive and conventional test data scored by the two-parameter model, correlations among θ estimates were very high, regardless of test length.

Both the one-and two-parameter hypothetical data illustrated the tendency of the Bayesian θ estimates to be regressed toward the mean. That is, the Bayesian scoring method provided lower θ estimates for scores above the mean and higher θ estimates for scores below the mean, in comparison to the two M-L scoring methods. This trend continued in the three-parameter data, although both rank-order and product-moment correlations remained high, as in the former two analyses. This result, however, has implications for the use of the Bayesian scoring method in any applied situation in which the absolute, as opposed to relative, level of the θ estimates is of importance. Since the Bayesian scoring method tends to restrict the range of θ estimates by imposing a normal distribution on them, θ estimates beyond ± 2.0 will rarely be obtained. The result is likely to be a tendency for this scoring method to fail to identify and/or to distinguish accurately among testees with extreme θ estimates.

The dissimilarities among the three scoring methods became most evident when the data were scored using the three-parameter model. The major dissimilarity, evident in all three data sets, was between the Bayesian and M-L logistic methods. In the adaptive test data, the Bayesian scoring method produced θ

estimates which had lowest correlations with one of the two M-L methods at all test lengths. For conventional tests of less than 15 items and for adaptive tests at all the lengths used in this study, these differences were substantial, indicating markedly different orderings of individuals, as in the hypothetical test data.

The three-parameter data also illustrated two other trends. First, the hypothetical test data showed a tendency toward lower θ estimates when the c parameter was included in scoring. A second, and more practically troublesome, trend was the tendency toward more convergence failures with the three-parameter data. This result was obvious in both the hypothetical test data and the live-testing data. The tendency toward convergence failures for the M-L scoring methods was most obvious in the conventional test; the number of convergence failures in the adaptive test was considerably less than in the conventional test when number of items was equal. This occurred because adaptive tests tend to locate for each testee the region of the item pool in which the testee will answer about half of the items correctly and half incorrectly. Thus, except for the rare individual for whom the adaptive test item pool is completely inappropriate in difficulty, adaptive tests will result in response patterns that are more likely scorable by M-L methods. This is not true of fixed-item peaked conventional tests, which must be targeted for a specific population θ level and which may be too easy or too difficult for substantial numbers of testees, resulting in response patterns not scorable by M-L methods.

Choosing a Scoring Method

These data show that in an adaptive test or in a situation in which a short conventional test is being administered, the choice of one of the ICC-based methods over another may have an impact on the ranking of the students in a course of training. For these situations, it is important that educators choose a scoring method most aligned to their philosophy of grading. To determine the "correct" scoring method to use, the underlying philosophies of the different scoring methods may be viewed by examining the relationship of the scores obtained from a particular method to the ICC response model underlying the test.

This can be illustrated with the hypothetical test used in the example of the two-parameter model, which was borrowed, in part, from Samejima (1969). Because the item parameters for this test were known, the way in which each scoring method depends on the item difficulty and discrimination parameters of the items answered by the testees may be examined. From inspection of Table 3 for the two-parameter data, it can be seen that the Bayesian strategy gave results most similar to a number-correct scoring strategy, since it ordered individuals almost perfectly with respect to number correct. However, higher rankings resulted with the Bayesian scoring method for individuals correctly answering more difficult (high b) and more discriminating (high a) items. A disadvantage of this scoring approach, however, is that more weight is given to the early items in the test.

The M-L normal rankings can be characterized as being dependent upon both the a and b parameters, but the dependence is less easily described than that of the Bayesian strategy. The M-L normal estimates tended to reward

correct answers to difficult items or correct answers to more discriminating items and to penalize inconsistent response patterns (that is, incorrect answers to easy items and correct answers to difficult items). The M-L logistic rankings for this response model were independent of the difficulty of the items answered correctly or incorrectly. As pointed out earlier, rankings were totally dependent on the discriminatory power of the items answered incorrectly by the individual (see Samejima, 1969, for the theoretical rationale).

It appears, therefore, that under the two-parameter response model, the M-L normal scoring method allows the most freedom from number-correct scoring and makes the most use of the parameter values of the items. If educators feel that this "philosophy" is in accord with their own, then it is the one that should be used; if it is not, one of the other scoring methods may serve better.

In addition to this "philosophy of scoring" approach, some of the other characteristics of the scoring methods should be considered. For instance, the Bayesian method allows the use of prior information in obtaining an achievement level estimate. If this prior information is accurate, this might be an advantage for obtaining good θ estimates from a short test. Prior information is not useful for M-L estimation. But if available prior information is not correct, the M-L scoring methods will be more accurate than the Bayesian method.

One final difference between the Bayesian and M-L scoring methods may be of some importance to educators. When individuals are able to answer test questions correctly by guessing, as in a multiple-choice test, the three-parameter ICC response model is most appropriate for scoring the test responses. Using this response model, M-L scoring methods will fail to converge on a unique θ estimate in some cases. For conventional test response data (Table 6), the percentage of such failures remained rather high under both M-L scoring methods (at least 5%) until more than 27 items had been administered. At no test length did all cases converge in the conventional test data.

The adaptive testing procedure fared better in this respect (Table 7). After the adaptive administration of only 9 items, neither M-L scoring method failed to obtain θ estimates in more than 3% of the cases. Further, all response patterns resulted in convergent θ estimates at all test lengths greater than 33 items.

These results suggest that an educator might take two courses of action to avoid the estimation failures of M-L scoring methods. One approach is to use a Bayesian scoring method, but with cognizance of its tendency to regress all θ estimates toward the mean. The other solution, of course, is to use an adaptive testing procedure in conjunction with either M-L scoring method.

In the final analysis, however, the choice of scoring method should be based on the validity of scoring methods in the prediction of external criteria. This study has demonstrated that, at least under the three-parameter ICC model, different scoring methods will provide different θ estimates. Given this knowledge, the question becomes one of studying the validity of the scores obtained from the different scoring methods with respect to relevant external criteria in order to determine whether the observed differences result in the differential predictability of criterion performance.

References

Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, January 1979.

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A047495).

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828).

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A046062).

Lindgren, Bernard W. Statistical theory. New York: Macmillan, 1976.

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

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (NTIS No. AD A022964).

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

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

Reckase, M. D. Ability estimation and item calibration using the one- and three-parameter logistic models: A comparative study (Research Report 77-1). Columbia: University of Missouri, Educational Psychology Department, Tailored Testing Research Laboratory, November 1977.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement, 1969, 34 (4, Pt. 2, Monograph No. 17).

Sympson, J. B. Estimation of latent trait status in adaptive testing procedures. In D. J. Weiss (Ed.), Applications of computerized adaptive testing (Research Report 77-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977. (NTIS No. AD A038114).

Urry, V. W. A five-year quest: Is computerized adaptive testing feasible? In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Research and Development Center, PS-75-6). Washington, DC: U.S. Government Printing Office, 1975, pp. 97-102. (Superintendent of Documents Stock No. 006-00940-9).

Weiss, D. J. Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1974. (NTIS No. AD A004270).

Appendix: Supplementary Tables

Table A
Parameter Estimates for Items in the Conventional Test

Item No.	No. Testees	*a*	*b*	*c*
3060	1323	.86	-1.31	.29
3067	1217	1.07	-.76	.21
3065	1324	1.17	-1.66	.39
3056	1134	.71	.89	.26
3063	1084	.91	1.51	.37
3073	1314	1.43	-1.57	.31
3058	1283	1.05	-.43	.44
3274	1274	.85	-1.05	.26
3271	1166	.95	1.32	.30
3055	1265	1.71	-.65	.24
3072	1177	1.02	.65	.32
3057	1285	1.20	-1.35	.26
3064	1287	.94	.86	.24
3069	1247	.88	-.01	.48
3054	1258	1.29	-.93	.31
3066	1057	1.05	.53	.31
3268	1211	.97	-.28	.18
3267	1285	1.02	-1.22	.23
3272	1274	1.06	-.81	.37
3070	1252	.95	-1.28	.22
3008	891	.96	-1.75	.18
3019	782	1.31	.29	.29
3062	1215	1.47	.43	.30
3061	1078	.85	1.57	.30
3262	1275	.81	.47	.45
3263	1092	.99	2.29	.53
3447	1266	1.18	.93	.32
3443	1264	1.07	-1.64	.37
3438	1095	.70	.21	.27
3448	1294	1.40	.73	.30
3435	1258	.83	-.61	.42
3439	1091	1.36	.64	.32
3436	1018	1.12	1.59	.41
3449	1138	.91	1.26	.14
3440	957	1.52	2.00	.30
3437	1147	1.95	.66	.28
3427	773	.92	1.51	.26
3445	1282	1.19	.44	.34
3444	1139	.88	.78	.38

Table B
Item Number, Number of Testees in Parameterization Group, Discrimination (a), Difficulty (b), and Guessing (c) Parameters for Items in the Stradaptive Item Pool

Item	N	a	b	c
Stratum 9 (15 items)				
3209	740	2.50	2.29	.29
3417	539	2.50	3.00	.35
3033	328	1.54	2.44	.35
3440	957	1.52	2.00	.30
3251	523	2.50	2.39	.35
3406	519	1.31	2.48	.35
3045	680	1.02	2.48	.27
3242	613	.94	2.40	.35
3407	564	1.02	2.41	.29
3263	1092	.99	2.29	.35
3241	756	.91	2.09	.17
3414	368	.88	2.29	.32
3402	401	.83	2.44	.35
3247	718	.82	2.42	.35
3228	396	.67	2.49	.31
Mean		1.33	2.39	.32
Stratum 8 (20 items)				
3409	602	2.50	1.28	.00
3234	220	2.50	1.73	.00
3018	953	.89	1.25	.35
3204	505	1.14	1.66	.35
3422	589	1.47	1.50	.35
3411	767	1.36	1.23	.35
3250	373	.91	1.94	.29
3206	410	.74	1.51	.21
3410	427	1.30	1.34	.31
3429	780	1.25	1.24	.28
3419	342	1.23	1.48	.25
3421	750	1.17	1.15	.35
3436	1018	1.12	1.59	.35
3271	1166	.95	1.32	.30
3061	1078	.95	1.57	.30
3427	773	.92	1.51	.26
3449	1138	.91	1.26	.14
3063	1084	.91	1.51	.35
3074	671	.84	1.79	.35
3420	541	.68	1.62	.35
Stratum 7 (20 items)				
3408	451	2.50	1.05	.31
3437	1147	1.95	.66	.28
3258	911	1.24	.81	.35
3432	595	1.72	.67	.35
3048	589	1.35	.66	.33
3413	832	1.40	.76	.35
3448	1294	1.40	.73	.30
3439	1091	1.36	.64	.32
3219	520	1.23	.62	.21
3072	1177	1.02	.65	.32
3277	892	1.00	1.04	.35
3035	772	.90	.68	.28
3433	657	1.35	.86	.30
3447	1266	1.18	.93	.32
3064	1287	.94	.86	.24
3230	895	.90	.87	.35
3444	1139	.88	.78	.35
3012	653	.75	.80	.35
3260	877	.71	.84	.28
3056	1139	.71	.89	.26
Mean		1.22	.79	.31
Stratum 6 (19 items)				
3047	608	1.66	.44	.29
3079	952	1.61	.27	.35
3213	900	.93	.52	.35
3041	716	1.51	.23	.35
3062	1215	1.47	.43	.30
3405	770	1.40	.55	.32
3445	1282	1.19	.44	.34
3218	500	.82	.58	.12
3019	782	1.31	.29	.29
3207	915	.70	.46	.28
3431	780	.70	.28	.34
3000	844	1.24	.52	.35
3046	626	1.18	.24	.22
3042	626	1.15	.37	.27
3050	713	1.13	.35	.18
3066	1057	1.05	.53	.31
3034	639	1.01	.37	.28
3262	1275	.81	.47	.35
3438	1095	.70	.21	.27
Mean		1.14	.40	.29
Stratum 5 (15 items)				
3282	1037	2.06	-.02	.35
3220	896	1.79	-.03	.26
3005	831	1.43	.11	.35
3425	649	1.36	.17	.23
3039	908	1.12	.12	.00
3214	809	1.12	.03	.23
3412	664	1.12	.19	.35
3051	752	1.29	.21	.28
3279	969	.99	.01	.28
3403	626	.99	.18	.19
3069	1247	.88	-.01	.35
3211	628	.88	.01	.13
3002	929	.82	.13	.14
3426	870	.68	.07	.22
3423	682	.66	.16	.27
Mean		1.15	.09	.24
Stratum 4 (13 items)				
3256	649	2.31	-.33	.26
3430	903	1.15	-.30	.29
3031	851	1.47	-.33	.35
3254	653	3.38	-.17	.22
3237	895	1.54	-.37	.18
3404	897	.65	-.29	.35
3244	854	1.35	-.44	.23
3058	1283	1.05	-.43	.35
3240	702	.98	-.28	.15
3268	1211	.97	-.28	.18
3208	850	.76	-.16	.12
3006	676	.77	-.37	.33
3259	879	.69	-.41	.20
Mean		1.23	-.32	.25
Stratum 3 (19 items)				
3021	906	1.96	-.49	.21
3217	893	1.06	-.48	.14
3038	951	1.71	-.93	.00
3055	1265	1.71	-.65	.24
3215	887	1.59	-.82	.23
Stratum 3, cont.				
3011	864	1.32	-.86	.20
3435	1258	.83	-.61	.35
3216	809	1.27	-.62	.18
3054	1258	1.29	-.93	.31
3221	938	1.25	-.52	.17
3049	814	1.15	-.71	.18
3255	657	1.14	-.72	.26
3067	1217	1.07	-.76	.21
3246	656	1.10	-.72	.28
3022	620	1.01	-.48	.30
3272	1274	1.06	-.81	.35
3017	950	.99	-.58	.16
3076	1054	.94	-.73	.21
3224	869	.80	-.50	.37
Mean		1.22	-.68	.22
Stratum 2 (20 items)				
3023	667	2.40	-1.15	.35
3202	922	1.81	-.99	.21
3415	915	.85	-.96	.35
3245	885	1.34	-.96	.21
3236	667	1.26	-1.20	.33
3020	915	1.23	-1.28	.17
3028	677	1.12	-1.26	.35
3226	941	1.09	-.98	.20
3210	895	1.04	-1.22	.35
3239	960	1.04	-1.13	.21
3013	880	1.00	-.97	.35
3267	1285	1.02	-1.22	.23
3257	928	.98	-1.02	.25
3070	1252	.95	-1.28	.22
3036	872	.92	-1.18	.16
3014	907	.86	-1.24	.14
3060	1323	.86	-1.31	.29
3274	1274	.85	-1.05	.26
3238	837	.82	-1.06	.21
3032	857	.77	-1.06	.27
Mean		1.11	-1.13	.26
Stratum 1 (17 items)				
3077	1053	2.50	-1.39	.20
3027	667	1.67	-1.38	.35
3443	1264	1.07	-1.64	.35
3249	910	.91	-1.69	.17
3428	899	.90	-1.56	.35
3073	1314	1.43	-1.57	.31
3205	908	1.25	-1.53	.19
3078	1060	1.24	-1.65	.35
3057	1285	1.20	-1.35	.26
3065	1324	1.17	-1.66	.35
3235	906	1.15	-1.40	.28
3029	957	1.13	-1.50	.28
3201	902	1.07	-1.34	.23
3008	891	.96	-1.75	.18
3252	898	.79	-1.77	.35
3003	914	.96	-1.76	.34
3044	913	.87	-1.42	.15
Mean		1.19	-1.55	.28

Table C

Correlations Between Achievement Level Estimates from Three Scoring Methods at Various Test Lengths for Conventional Test Data Scored by One-, Two- and Three-Parameter Models

(N=200[a])

Number of Items	One-Parameter Model				Two-Parameter Model				Three-Parameter Model					
		MLL vs. MLN	MLL vs. Bayes	MLN vs. Bayes		MLL vs. MLN	MLL vs. Bayes	MLN vs. Bayes	MLL vs. MLN		MLL vs. Bayes		MLN vs. Bayes	
	N				N				N	r	N	r	N	r
3	75	.9955	.9741	.9832	75	.9958	.9629	.9749	70	.8957	71	.7917	70	.9852
6	146	.9831	.9867	.9964	146	.9874	.9776	.9876	140	.9463	141	.8994	142	.9881
9	167	.9847	.9892	.9957	167	.9864	.9858	.9921	167	.9890	167	.9797	167	.9892
12	174	.9858	.9888	.9959	174	.9868	.9859	.9920	174	.9619	174	.9527	174	.9828
15	181	.9893	.9904	.9963	181	.9889	.9881	.9939	181	.9953	181	.9792	181	.9846
18	184	.9920	.9926	.9967	184	.9918	.9910	.9951	184	.9960	184	.9800	184	.9822
21	184	.9912	.9920	.9966	184	.9916	.9912	.9947	184	.9950	184	.9877	184	.9882
24	188	.9921	.9935	.9960	188	.9924	.9922	.9947	188	.9953	188	.9901	188	.9905
27	191	.9918	.9905	.9957	191	.9919	.9890	.9944	191	.9953	191	.9900	191	.9909
30	192	.9928	.9912	.9959	192	.9927	.9897	.9950	192	.9953	192	.9932	192	.9961
33	192	.9935	.9915	.9961	192	.9934	.9897	.9952	192	.9957	192	.9938	192	.9963
36	198	.9946	.9925	.9958	198	.9945	.9907	.9948	198	.9961	198	.9943	198	.9954
39	198	.9954	.9928	.9958	198	.9951	.9911	.9948	198	.9967	198	.9948	198	.9960

[a]Differences in N from 200 represent nonconvergent cases in M-L scoring.

Table D

Correlations Between Achievement Level Estimates from Three Scoring Methods at Various Test Lengths for Adaptive Test Data Scored by One-, Two-, and Three-Parameter Models

Number of Items	One-Parameter Model						Two-Parameter Model						Three-Parameter Model					
	MLL vs. MLN		MLL vs. Bayes		MLN vs. Bayes		MLL vs. MLN		MLL vs. Bayes		MLN vs. Bayes		MLL vs. MLN		MLL vs. Bayes		MLN vs. Bayes	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*
3	159	.9992	159	.9927	159	.9960	159	.9854	159	.9854	159	.9871	108	.9997	141	.9334	124	.9456
6	188	.9995	188	.9978	188	.9988	188	.9923	188	.9923	188	.9933	174	.8628	177	.8444	182	.9563
9	198	.9998	198	.9986	198	.9991	198	.9946	198	.9946	198	.9950	189	.9970	189	.9315	192	.9479
12	185	.9997	185	.9984	185	.9989	185	.9920	185	.9920	188	.9937	182	.9228	182	.9243	184	.8832
15	169	.9997	169	.9989	169	.9995	169	.9923	169	.9923	169	.9941	166	.9382	166	.9599	168	.9154
18	143	.9997	143	.9990	143	.9995	143	.9930	143	.9930	143	.9944	150	.9478	140	.9520	142	.9226
21	127	.9995	127	.9989	127	.9996	127	.9924	127	.9924	127	.9946	123	.9883	123	.9617	125	.9711
24	108	.9995	108	.9990	108	.9998	108	.9915	108	.9915	108	.9946	108	.9846	108	.9201	108	.9485
27	97	.9995	97	.9990	97	.9998	97	.9914	97	.9914	97	.9951	96	.9880	96	.9534	96	.9702
30	83	.9995	83	.9990	83	.9998	83	.9909	93	.9909	83	.9944	81	.9830	82	.9536	81	.9769
33	75	.9994	75	.9990	75	.9998	75	.9910	75	.9910	75	.9946	74	.9827	75	.9566	74	.9766
36	67	.9994	67	.9989	67	.9998	67	.9900	67	.9900	67	.9948	67	.9810	67	.9624	67	.9815
39	60	.9993	59	.9989	60	.9998	60	.9900	60	.9900	60	.9946	60	.9804	60	.9648	60	.9822
42	56	.9995	45	.9990	56	.9998	56	.9907	56	.9807	56	.9946	56	.9800	56	.9656	56	.9823
45	51	.9994	51	.9990	51	.9998	51	.9906	51	.9906	51	.9946	51	.9806	51	.9669	51	.9828
48	47	.9994	47	.9988	47	.9997	47	.9892	47	.9892	47	.9933	47	.9765	47	.9598	47	.9784

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE
HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research
Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION &
TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Head, Research, Development, and Studies
(OP102X)
Office of the Chief of Naval Operations
Washington, DC 20370

1 Scientific Advisor to the Chief of
Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH CANTER
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

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

1 MCDEC
Quantico Marine Corps Base
Quantico, VA 22134

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

CoastGuard

1 MR. JOSEPH J. COWAN, CHIEF
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Vern W. Urry
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

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

Non Govt

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

1 Dr. Erling B. Andersen
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Werner Birke
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Progr
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. M. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Psychology
University of Illinois
Champaign, IL 61820

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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

1 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

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

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

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

Final Report: Bias-Free Computerized Testing. March 1979.

79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.

79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.

78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.

78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.

78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.

78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.

78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.

77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.

77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.

77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.

77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.

77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.

77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.

77-1. Applications of Computerized Adaptive Testing. March 1977.

Final Report: Computerized Ability Testing, 1972-75. April 1976.

76-5. Effects of Item Characteristics on Test Fairness. December 1976.

76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.

76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.

76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.

76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.

75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.

75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.

75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.

75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.

75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.

75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.

74-5. Strategies of Adaptive Ability Measurement. December 1974.

74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.

74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.

74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.

74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.

73-3. The Stratified Adaptive Computerized Ability Test. September 1973.

73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.

73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

EFFECT OF POINT-IN-TIME IN INSTRUCTION ON THE MEASUREMENT OF ACHIEVEMENT

G. Gage Kingsbury
and
David J. Weiss

RESEARCH REPORT 79-4
AUGUST 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Navy Personnel Research and Development Center, Army Research Institute, Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, and the Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Effect of Point-in-Time in Instruction on the Measurement of Achievement		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* G. Gage Kingsbury and David J. Weiss		8. CONTRACT OR GRANT NUMBER*(s)* N00014-76-C-0627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.: NR150-389
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE August 1979 13. NUMBER OF PAGES 27
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 in 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

This research was supported by funds from the Navy Personnel Research and Development Center, Army Research Institute, Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, and the Office of Naval Research, and monitored by the Office of Naval Research.

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

achievement testing	pretest-test measurement	change scores
latent trait test theory	test-posttest measurement	adaptive testing
item response theory	measurement of change	tailored testing
item characteristic curve theory	measurement of growth	

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

Item characteristic curve (ICC) theory has potential for solving some of the problems inherent in the pretest-test and test-posttest paradigms for measuring change in achievement levels. However, if achievement tests given at different points in the course of instruction tap different achievement dimensions, the use of ICC approaches and/or change scores from these tests is not desirable. This problem is investigated in two studies designed to determine whether or not achievement tests administered at different times

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

during a sequence of instruction actually measure the same achievement dimensions.

To investigate possible changes in dimensionality between different points in instruction, aspects of the dimensionality of achievement test data were examined prior to instruction, at the peak of instruction, and up to a month following the peak of instruction. Data used were conventional and adaptive achievement test data administered to students in a general biology course at the University of Minnesota.

Results raised questions about the utility of the pretest-test paradigm for measuring change in achievement levels, since a comparison of ICC parameter estimates indicated that a change in the dimensionality of achievement had occurred within the short (4-week) period of instruction. This change was also observed using a factor analytic comparison.

Use of the test-posttest paradigm to measure retention was supported, since a regression comparison of students' achievement level estimates did not indicate any significant change in the achievement metric up to 1 month after the peak of instruction. The significance of this result for the use of adaptive testing technology in measuring achievement is described.

Implications of these studies and the use of ICC theory in the measurement of achievement, as well as some potential limitations in terms of generalizability of these results, are discussed.

CONTENTS

Acknowledgments

Adaptive and conventional test data utilized in this report were obtained from volunteer students in Biology 1-011, General Biology, at the University of Minnesota during fall quarters 1976 and 1977 and winter quarter 1977; appreciation is extended to these students for their participation in this research. The cooperation of Kathy Swart and Norman Kerr of the General Biology staff in providing access to the students is also deeply appreciated.

Technical Editor: Barbara Leslie Camm

Effect of Point-in-Time in Instruction on the Measurement of Achievement

The measurement of achievement is a considerably more complex problem than the measurement of ability. Whereas ability levels develop over long periods of time and remain relatively stable with exposure to different environments, achievement levels characteristically result from exposure to specific instructional or training environments. Although these instructional environments may span relatively long periods of time, such as the 2- to 4-year periods required for training in many skilled fields, instructional decisions are more typically made on the basis of instructional periods of several weeks or a few months. In the extreme case, as in computer-assisted instruction, the instructional environment designed to modify individuals' achievement levels may be as short as a few minutes. Thus, achievement is a dynamic variable which may change over very short time intervals.

The Pretest-Test Paradigm

Because achievement levels should be sensitive to instruction, it may be desirable to measure achievement (1) prior to instruction, (2) at the end (or peak) of instruction, and (3) some time after the completion of instruction. The measurement of achievement prior to instruction is accomplished by means of pretests, designed to determine a student's level of achievement before exposure to the instructional environment. To determine whether instruction has had an effect, a pretest-test paradigm may be used, in which group or individual gain scores are computed to demonstrate the effects of the instruction.

The pretest-test paradigm for measuring achievement has at least two major problems. First, both individual and group change scores have been shown to be highly unreliable (Cronbach & Furby, 1970) unless the pretest and peak-of-instruction test measurements are each extremely precise. Second, it is necessary to administer two tests covering the same material to students. If the same test is administered prior to instruction and at the peak of instruction, students may not be motivated to respond optimally to both tests; thus, precision of measurement on the tests may be lowered. At the same time, students may remember test items from the pretest, find the answers to those items in the course materials, and then perform better on the second test than they would have performed had they not seen the test items prior to instruction.

On the other hand, if different tests are used prior to instruction and at the peak of instruction, the serious problem of developing parallel tests arises. This is compounded by the necessity to obtain very precise measurements, which may result in pretests and peak-of-instruction tests measuring different aspects of the achievement variable in the quest for highly precise measurements.

A potential solution to these problems in the pretest-test paradigm lies in the application of techniques of item characteristic curve (ICC) theory

(Lord & Novick, 1968) to the measurement of achievement. Achievement test items calibrated using ICC theory (e.g., Bejar, Weiss, & Kingsbury, 1977) will all be on the same metric. Thus, selection of ICC-calibrated items from the same item pool to constitute pretests and peak-of-instruction tests will, in theory, eliminate the need for the construction of parallel tests. In addition, placing all the items on the same metric by using ICC item parameters will eliminate the need to repeat the same items at the two testings, since (again in theory) any subset of items from the precalibrated pool will measure the same variable as any other subset of items. Thus, items for pretests and for peak-of-instruction tests can be selected from the ICC-calibrated pool on the basis of content considerations resulting in effectively parallel measurements.

ICC theory can also be applied to the pretest-test paradigm of achievement measurement through the use of adaptive testing to increase the precision of the achievement measurements, thus possibly permitting the use of individual or group gain scores. Research by Bejar and Weiss (Bejar & Weiss, 1978; Bejar, Weiss, & Gialluca, 1977) in an achievement testing context shows that ICC-based adaptive tests produce measurements which are considerably more precise than those of conventional achievement tests, supporting similar findings in the ability-testing literature (e.g., McBride & Weiss, 1976; Vale, 1975; Vale & Weiss, 1975).

Before ICC theory can be applied in the pretest-test paradigm of achievement measurement, however, it must be demonstrated that data obtained in this paradigm meet the assumptions of the theory. Specifically, since most ICC-based techniques require unidimensionality, it must be shown that both pretests and peak-of-instruction tests are essentially unidimensional or that, in general, the dimensionality of the two tests is the same. In addition, it must be demonstrated that the latent space in the two tests does not change. That is, even though both the pretest and the peak-of-instruction test are unidimensional, it is possible that they are measuring achievement on different dimensions. If this is the case, item parameters estimated at one point in instruction would not be usable at the other point in instruction.

The Test-Posttest Paradigm

Just as measured achievement is expected to change in level from pretest to peak of instruction, it is also expected that it should remain stable for some time after instruction. Thus, it is appropriate to investigate whether measured achievement levels deteriorate over short or long periods of time in order to determine the permanency of the instructional effect demonstrated by the pretest-test data. Such a demonstration would require the test-posttest paradigm in which the peak-of-instruction test is followed at some point in time by the administration of a posttest.

Because the test-posttest paradigm for measuring constancy of achievement may be implemented with either the parallel tests approach or the repeated tests approach, it has exactly the same problems as the pretest-test paradigm. Similarly, the use of ICC theory and adaptive testing may be brought to bear on these problems if the peak-of-instruction and posttest data meet the requirements of these approaches. Thus, similar kinds of data must be generated to investigate the use of these approaches in an achievement context.

The use of adaptive testing in the measurement of achievement--whether at pretest, peak of instruction, or posttest--raises an additional problem which

requires investigation. To realize the potential gains in the measurement of achievement in increased precision (Bejar, Weiss, & Gialluca, 1977), higher validity (Bejar & Weiss, 1978), and shorter testing times (Brown & Weiss, 1977), adaptive tests should be administered by computer. In achievement environments with large numbers of students, there may not be sufficient numbers of adaptive testing terminals so that each student can be tested at the peak of his or her instruction. Thus, it may be necessary for students to have their achievement measured at some point beyond the peak of instruction. A similar situation exists in self-paced instructional environments, where students may not take achievement tests exactly at the peak of instruction due to procrastination or influences beyond their control (e.g., unavailability of equipment). In both cases, it is an important question whether achievement measured after the peak of instruction is measured on the same dimension as achievement measured at the peak of instruction.

Objectives

The studies reported below were designed to investigate several questions relevant to the implementation of ICC theory in pretest-test and test-posttest paradigms for measuring achievement. The data also have some bearing on the practical questions involved in the use of adaptive testing in measuring achievement within the realistic constraints of instructional environments.

STUDY 1: RELATIONSHIP OF TEST CHARACTERISTICS PRIOR TO INSTRUCTION AND AT PEAK OF INSTRUCTION

This study was designed to investigate two questions concerning test characteristics of a test used to measure achievement at the peak of instruction when it was applied to a population of testees measured prior to instruction:

1. Are ICC item parameters estimated from data obtained prior to instruction quantitatively equivalent to parameters estimated at the peak of instruction? This question is concerned with whether the ICC metric maintains its interval properties during the course of instruction.

2. Do tests used to measure achievement prior to instruction (i.e., pretests) measure attributes from the same latent space as tests used to measure achievement at the peak of instruction? This question is concerned with whether the responses of the two populations (pretest versus test) can be described by a common latent space.

If the responses to both of these questions are affirmative, the results may be taken as support for the pretest-test paradigm. These results would also have implications for the power of the unidimensional ICC model for measuring achievement during the course of instruction. If major differences are found in the characteristics of the tests used to measure achievement prior to and at the peak of instruction, the foundation of the pretest-test paradigm for measuring achievement would be weakened in many applications and the use of the ICC model to measure growth in achievement would be limited.

Method

Test Data

Prior to instruction. The testing sessions that provided the prior-to-instruction data were classroom examinations administered on the first day of class during the fall academic quarter of 1977 to all students attending class for Biology 1-011, General Biology, at the University of Minnesota. (For a description of the course and testing procedures, see Bejar, Weiss, & Kingsbury, 1977.) Data were obtained from 1,294 students. The test administered at this time consisted of 40 multiple-choice items sampled from all of the 7 content areas covered in the course; these items were taken from a larger pool of items developed for this course.

Peak of instruction. The peak-of-instruction test data were obtained from two sources so that two different types of questions could be answered. The first question concerned whether item parameter estimates obtained from prior-to-instruction testing were similar to estimates obtained for the same items at peak-of-instruction testing. Peak-of-instruction parameter estimates were used that were obtained from test data supplied by individuals enrolled in the same course during five earlier academic quarters, since it would be inappropriate to administer the same items twice to the same individuals. These calibration samples averaged between 700 and 1,000 students. (For the exact number of subjects responding to each of the items for calibration, see Kingsbury & Weiss, 1979, p. 26.)

The second question concerned whether the factor structure underlying students' responses changed as a function of instruction. To answer this question, student reponse data were used which were collected on a 55-item midquarter examination administered 4 weeks after the pretest, as one of the course requirements, to approximately the same group of students who took the pretest. Approximately 1,200 students completed the 55-item midquarter examination. Each student was required to omit 5 items in the examination; consequently, data for each item were based on about 1,000 students.

Item Parameter Analysis

Item parameterization. Estimation of ICC item parameters for the peak-of-instruction data is described in detail in Bejar, Weiss, and Kingsbury (1977). In brief, a computer program developed by Urry (1976) was used to fit a three-parameter logistic ogive for each item administered to the testees. Items were rejected by the parameter estimation program if they failed to reach certain minimal standards with respect to their discrimination value (a) and lower asymptote (c). Thus, values for the index of discriminatory power (a), (b), and probability of attaining a correct answer with no knowledge of the subject (c) were obtained for each item that surpassed the minimum standards. Specifically, an item was rejected if during the first stage of the parameter estimation process the value of its a parameter estimate was less than .80 or the value of its c parameter estimate was greater than .30.

This procedure was applied separately to both the prior-to-instruction data and the peak-of-instruction data. The results, for each of the 40 items administered at both points in instruction, were two comparable estimates of the ICC parameter values, varying only because of sample fluctuation and the difference in the instructional level of the two groups.

Comparison of item parameter estimates. If the latent space is constant, the parameter estimates from the prior-to-instruction and the peak-of-instruction groups should differ no more than the estimates obtained from two samples at the same level of instruction. To provide a basis of comparison for the prior-to-instruction versus peak-of-instruction correlations, ICC item parameter estimates were computed from two groups of students in the same course during two earlier quarters who answered a comparable group of items at the peak of instruction (the peak/peak group). The peak/peak data were partially reported before by Bejar, Weiss, and Kingsbury (1977), who reported the peak/peak correlations for the a and b parameter estimates for 18 items obtained from responses of approximately 900 testees in each sample. Appendix Table A shows item numbers and parameter estimates of the items used to investigate sampling variation in parameter estimates obtained from the two groups at the peak of instruction. Also in this table are the times of administration of the items to the students.

To compare ICC parameter estimates obtained prior to instruction with those obtained at the peak of instruction, Pearson product-moment correlations were computed between item parameter estimates obtained at the two time periods (prior/peak correlations) separately for the a, b, and c parameters. The three correlations obtained were also computed in the two samples which were at the peak of instruction (peak/peak correlations). For each of the three parameters, the prior/peak and peak/peak correlations should differ only to the extent that the individuals differed in their test performance when the group was tested prior to instruction compared with their performance when tested at the peak of instruction, as reflected in the ICC item parameters.

To determine whether the prior/peak correlations differed significantly from the peak/peak correlations, Fisher's z'-transformation was applied to the prior/peak correlations and a confidence interval was constructed around each correlation (Neter & Wasserman, 1974). If these intervals included the observed peak/peak correlations, the hypothesis that the obtained correlations might come from the same population could not be rejected. If a confidence interval around the prior/peak correlation did not include the value of the observed peak/peak correlation, it could be concluded that the differences between the observed correlations were probably not due to sampling fluctuation. If the peak/peak correlation fell above the upper limit of the prior/peak confidence interval, this would imply that the ICC parameters were not invariant between the prior-to-instruction sample and the peak-of-instruction sample. This variability of parameter values would indicate that the two samples reflected different populations and that the ICC parameters estimated in the peak-of-instruction population were not sufficient to describe the responses of individuals in the prior-to-instruction population.

Factor Structure

Of the items administered in the pretest, 21 were sampled from the content areas taught in the first portion of the course; these content areas were then tested on the first midquarter examination, which was administered later in the course. The items were used to investigate the factor structure prior to instruction.

Twenty-one items tapping the same content areas were chosen arbitrarily from the first midquarter examination, which was administered to the same individuals 4 weeks after the pretest. The students' responses to these items

were used to examine the factor structure underlying performance at the peak of instruction. Items administered at the pretest and at the first midquarter were sampled from the same content areas, but different items were used at the two points in time.

For each of these groups of items, the same procedure was followed to obtain the final factor structure. First, all student responses were scored "0" if incorrect or "1" if correct. Second, these recoded responses were used to obtain tetrachoric correlations among the items through the TETRACHORIC subroutine in the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1970). The two resultant correlation matrices were then factor analyzed using the FACTOR subroutine from the same statistical package. The final factor solutions were obtained using a principal axis solution; the initial communality estimates were the squared multiple correlations of each variable with all the other variables. The factor solutions, which were arbitrarily limited to five factors, were iterated until the differences in successive communality estimates were negligible. This procedure provided the final solutions.

The two final factor solutions were then compared for similarities and differences in terms of the number of salient factors, the strength of each factor, and the amount of variance in the item intercorrelations accounted for by the factor solutions. To the extent that observed differences between the two solutions were minor, it could be inferred that the underlying factors contributing to testee responses were the same prior to instruction and at the peak of instruction. To the extent that major discrepancies were observed between the solutions, it could be inferred that differences existed in the structure of achievement at the two points in instruction.

Results

Item Parameters

Parameter estimates for each of the 40 items administered in the prior-to-instruction achievement measure are shown in Table 1, along with parameter estimates for the same items obtained from groups of testees at the peak of instruction. From this table it can be seen that 14 of the items failed to meet the minimal standards of the estimation procedure when administered prior to instruction and 5 items failed when administered at the peak of instruction. Four items were rejected in both instances. After all of the rejected items were removed from consideration, 25 items remained for the correlational analysis.

The bivariate plots of a, b, and c parameter estimates from the two calibrations are shown in Figures 1, 2, and 3, respectively. It can be seen from these figures that the relationships between the sample estimates of the parameter values were weak, at best. Figure 1 shows a correlation of -.12 for the a parameter, indicating a slight tendency for high values of a at peak of instruction to be associated with low values prior to instruction.

The correlation of r=.64 for the b parameter data show a tendency for high values of b prior to instruction to be associated with high values at the peak of instruction. However, almost all the data points in Figure 2 are below the main diagonal, indicating a tendency for items to be more difficult prior to

Table 1
Parameter Estimates of Items Calibrated
Prior to Instruction and at the Peak of Instruction

Item Number	Prior to Instruction			Peak of Instruction		
	a	*b*	*c*	*a*	*b*	*c*
3035	1.21	2.53	.28	.90	.68	.28
3241	1.89	2.87	.48	.91	2.09	.17
3816	1.72	-.26	.51	--	--	--
4013	1.02	.79	.38	1.76	-1.88	.16
3809	1.12	.70	.35	1.27	-.61	.53
4010	1.25	.00	.43	.88	-1.82	.23
3817	--	--	--	--	--	--
3803	--	--	--	--	--	--
3210	--	--	--	1.04	-1.22	.40
3837	1.10	1.13	.29	1.09	-1.59	.25
3235	1.17	1.38	.49	1.15	-1.40	.28
3808	--	--	--	.99	-1.00	.30
4033	--	--	--	.90	2.23	.38
3812	.85	1.92	.33	.82	-.63	.13
3424	--	--	--	--	--	--
3821	--	--	--	.90	-.92	.43
3244	1.00	2.52	.47	1.35	-.44	.23
3013	.91	1.18	.38	1.00	-.97	.39
3065	1.55	.03	.46	1.17	-1.66	.39
3909	--	--	--	1.34	.77	.38
3922	.76	2.33	.28	.64	-.26	.30
3415	--	--	--	.85	-.96	.41
3428	--	--	--	.90	-1.56	.40
3067	.98	1.15	.33	1.07	-.76	.21
3272	.83	1.45	.40	1.06	-.81	.37
3908	--	--	--	1.15	.07	.31
3435	1.85	1.43	.39	.83	-.61	.42
4005	--	--	--	--	--	--
3426	1.03	2.68	.44	.68	.07	.22
3031	.75	2.54	.24	1.47	-.33	.39
4006	1.01	2.53	.47	.84	-.59	.16
3069	.85	1.11	.45	.88	-.01	.48
3211	--	--	--	.88	.01	.13
3905	--	--	--	.98	.35	.20
4015	.76	2.26	.31	2.03	-1.62	.12
3403	.93	1.38	.33	.99	.18	.19
3000	3.06	2.46	.21	1.24	.52	.36
3445	.73	2.50	.39	1.19	.44	.34
3218	2.44	2.18	.30	.82	.58	.12
4001	1.44	2.49	.29	1.47	-1.14	.13

Note. Missing values indicate that item was rejected by the item calibration procedure.

instruction than at the peak of instruction. The data in Table 1 indicate that prior to instruction only one of the item *b* values was negative (an easy item), but at peak of instruction more than half the items had negative *b* values.

Figure 1
ICC Discrimination (a) Parameter Values Estimated
Prior to Instruction and at Peak of Instruction (r=-.12)

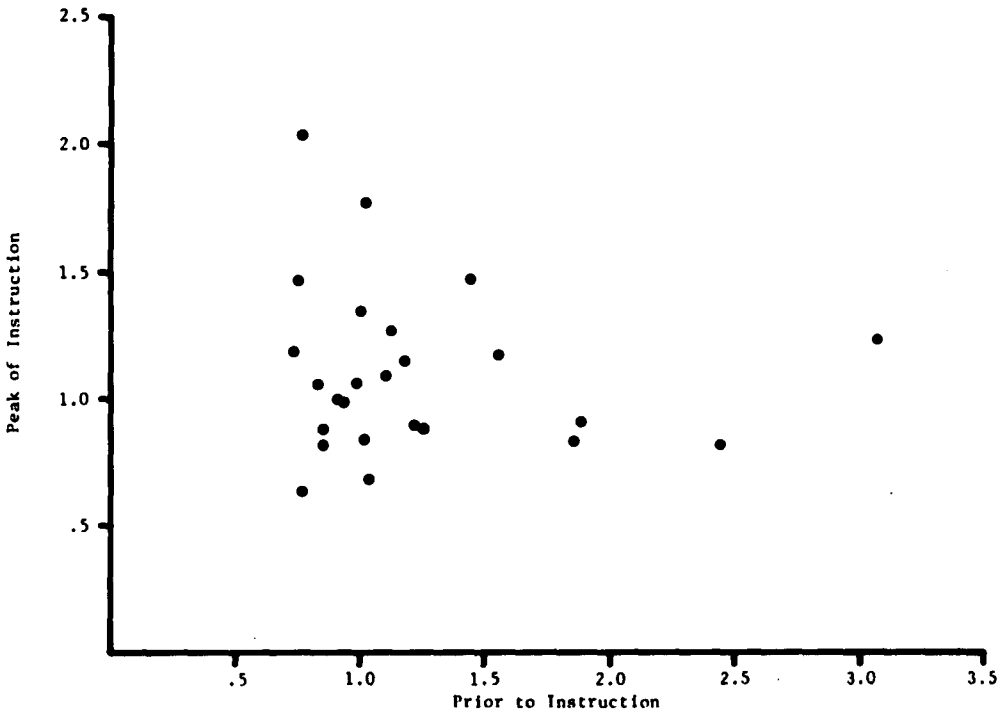


Figure 2
ICC Difficulty (b) Parameter Values Estimated
Prior to Instruction and at Peak of Instruction (r=.64)

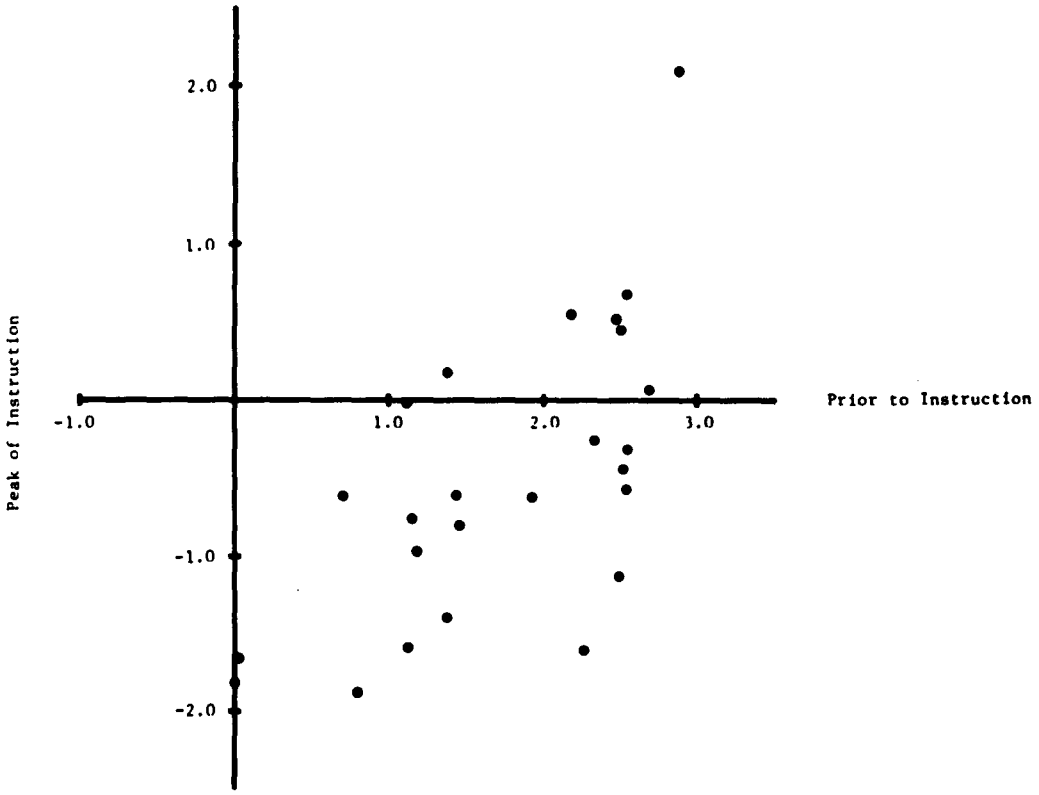
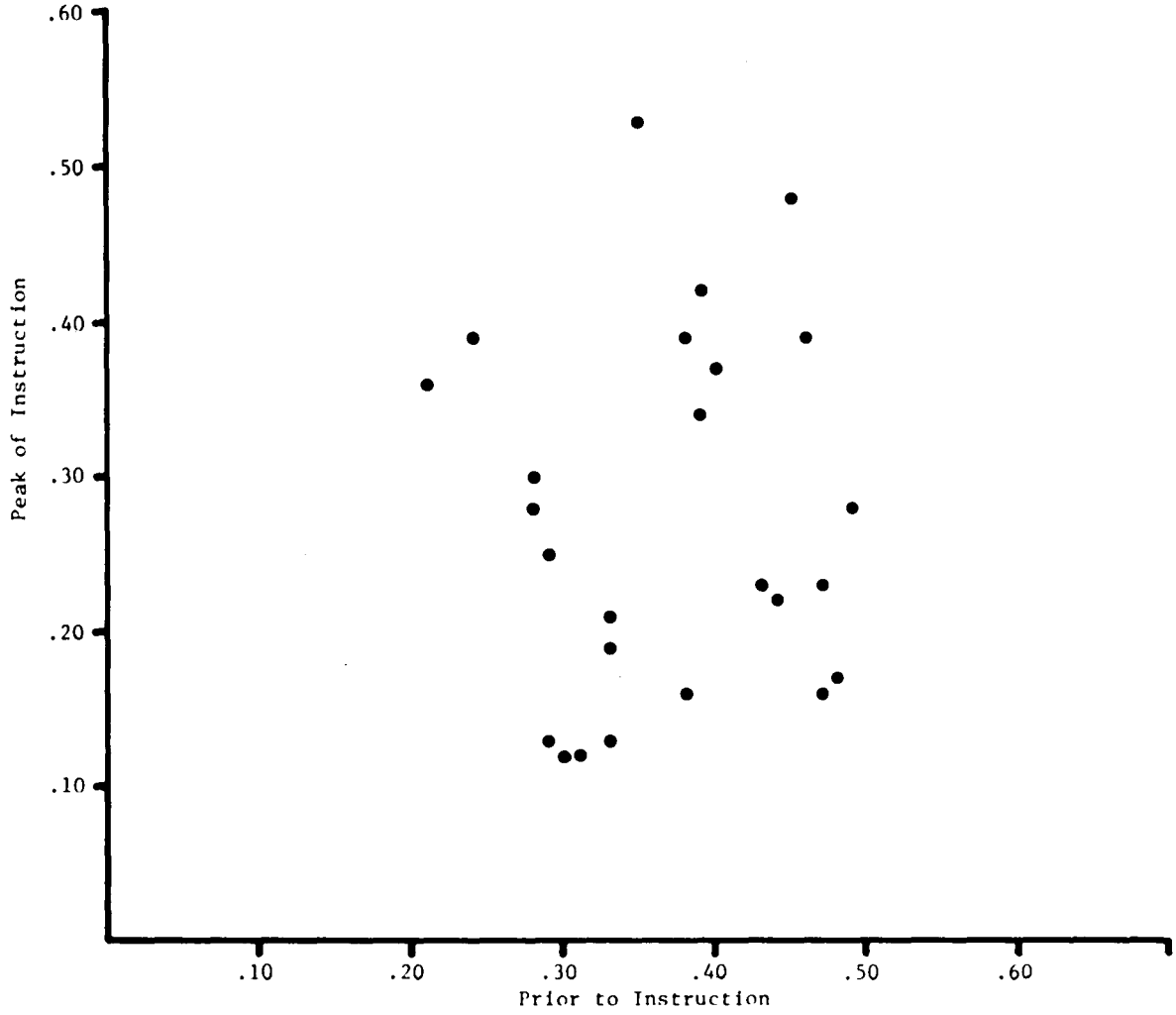


Figure 3
ICC Lower Asymptote (c) Parameter Values Estimated Prior to Instruction And at Peak of Instruction (r=.04)



The data in Figure 3 indicate essentially no relationship (r=.04) between the c parameters estimated at the two points in time. As expected, however, there was a general tendency for guessing parameter values to be higher prior to instruction than at the peak of instruction.

Table 2
Correlations Between Item Parameter Estimates for Prior/Peak and Peak/Peak Data

Data	a	b	c
Prior/Peak	-.12	.64	.04
Peak/Peak	.63	.96	.41

The prior/peak correlations (based on item parameter data in Table 1) are shown in Table 2, along with the peak/peak correlations (based on data in Appendix Table A) obtained from separate samples on other items drawn from the same

testing pool. Using the z'-transformation, 95% confidence intervals were computed for each of the prior/peak correlations with the following results:

1. For the index of item discrimination, a, the lower limit of the 95% confidence interval around the prior/peak correlation was -.50; the upper limit was .30. The peak/peak correlation for this parameter was .63, which was beyond the bounds of the confidence interval.

2. For the index of item difficulty, b, the limits of the 95% confidence interval were .31 and .81. The peak/peak correlation was .96, which was beyond the limits of the confidence interval.

3. For the index of the lower asymptote of the ICC, c, the limits of the 95% confidence interval were -.37 and .55. The peak/peak correlation, .41, fell within the bounds of the confidence interval. It should be noted that the correlation between the estimates of the c-parameter was quite low, even in comparable samples, accounting for less than 20% common variance.

Factor Structure

Item intercorrelation matrices for the two 21-item subsets are in Appendix Table B. Final values of the communality estimates are shown in Table 3. The

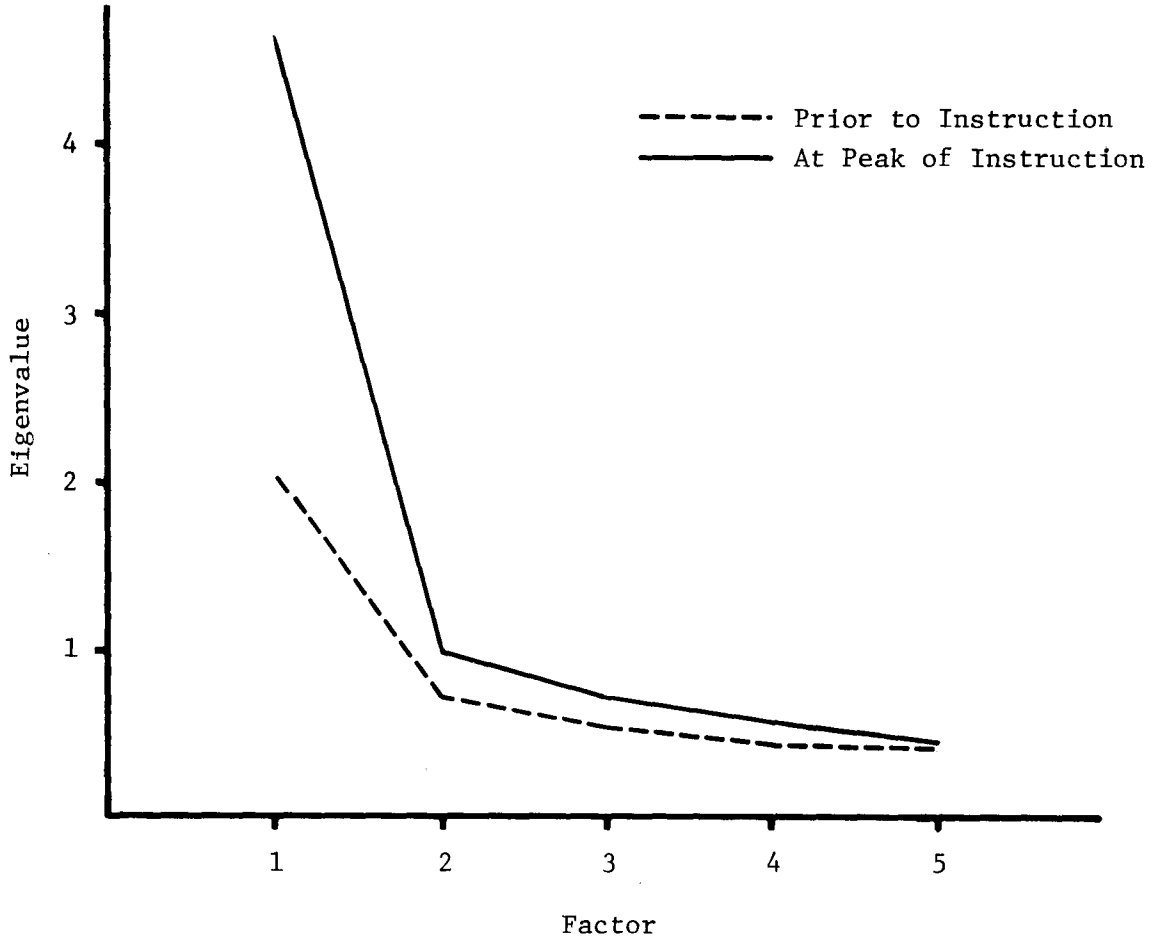
Table 3
Final Communality Estimates
Used in Factor Analysis of
Pretest and First Midquarter Examination

Item	Pretest	First Midquarter
1	.351	.929
2	.024	.166
3	.191	.233
4	.476	.140
5	.127	.960
6	.052	.411
7	.198	.134
8	.378	.218
9	.068	.279
10	.079	.495
11	.242	.345
12	.093	.250
13	.709	.198
14	.031	.451
15	.225	.346
16	.379	.487
17	.110	.192
18	.265	.188
19	.092	.553
20	.034	.365
21	.171	.114

mean communality estimate for the pretest items prior to instruction was .20; the mean communality estimate for the first midquarter items at the peak of instruction was .34. About one-third of the variance for the average test item at the peak of instruction was accounted for by the common factors, compared to one-fifth of the variance for the average item prior to instruction. Consequently, there was more unique variance in the items administered prior to instruction than in the items administered at the peak of instruction.

Eigenvalues for the final factor solutions are presented graphically in Figure 4 (factor loadings are in Appendix Table C). It can be seen that subsequent to the first factor, the eigenvalues for each factor were quite similar. For the first factor, though, the eigenvalue obtained from the items administered at the peak of instruction (4.67) was more than twice the value obtained from those administered prior to instruction (2.04). Thus, although both sets of items had a dominant first factor, in the peak-of-training data a stronger first factor was evident; in the prior-to-training data a weak single factor accounted for achievement.

Figure 4
Eigenvalues of Factors Obtained Prior to Instruction (Pretest) and at Peak of Instruction (First Midquarter Examination)



Conclusions

The correlational analysis indicated that the ICC parameters obtained from two samples differing in instruction were more discrepant than was to be expected from observing correlations among parameters obtained from samples at the same point in instruction. Further, for two of the ICC parameters--discrimination (a) and difficulty (b)--the prior/peak correlations were found to be significantly smaller than the peak/peak correlations. These findings imply that the latent space underlying testee responses changes enough between the beginning and the peak of instruction so that the test responses cannot be described by a single latent continuum; items change, not only with respect to their ability to differentiate testees at different levels of the trait continuum (discrimination), but also with respect to their relative difficulties. The findings imply that in order to describe test item responses obtained both prior to and at the peak of instruction with a single latent trait model, the unidimensional model that was considered adequate to describe performance at the peak of instruction would need to be expanded to a multidimensional model developed from both sets of data.

This conclusion is supported by the comparative factor analyses. These analyses showed a major difference in the strength of the first factor underlying testee responses in the few weeks of instruction between the pretest and the peak of instruction. Students' test item responses were not, to as great an extent, related to the first factor during the pretest. Again, the implication is that students were not responding to the same influences to the same degree on the pretest and on the midquarter examination.

These findings, taken as a whole, imply that the pretest-test paradigm may be invalid in some instances simply because the tests might not be tapping the same underlying achievement variable. This may account, in part, for the lack of reliability of change scores reported in many studies (e.g., Cronbach & Furby, 1970; Harris, 1963). It is suggested that the underlying factor structures in testee responses be explored whenever possible before importance is attributed to any pretest-test measure of change.

STUDY 2: STABILITY OF ACHIEVEMENT ESTIMATES AFTER THE PEAK OF INSTRUCTION

The stability of achievement estimates measured after the peak of instruction is important for at least two reasons. First, it is frequently necessary to measure the achievement levels of some individuals at a different time than others. Students occasionally miss examinations for a variety of reasons and, therefore, take the examination at a point which may not be at their peak of instruction. Where tests are given by computers (e.g., as in adaptive testing), there may not be sufficient terminal equipment available to test all students immediately at the peak of instruction. It is thus important to determine whether the passage of time after the completion of instruction affects achievement level estimates.

Measuring an individual after the peak of instruction is also a common problem in research studies attempting to measure retention. Similar to the pretest-test paradigm, the test-posttest paradigm used to measure reten-

tion assumes that the same achievement variable is being measured and that the passage of time does not change the nature of achievement. Thus, it is again relevant to determine whether achievement level estimates obtained after the peak of instruction are systematically related to those obtained at the peak of instruction.

If it is hypothesized that performance of individuals tested some time after the peak of instruction is a function of the same latent space that influences performance at the peak of instruction, several outcomes would be expected when individuals' achievement levels estimated from tests given at different times are compared. If the unidimensional latent space remains static with the passage of time, achievement estimates for individuals measured at different points in time after the peak of instruction should differ only as a linear function of the time of testing after the peak of instruction. For instance, if a group of individuals passed through a particular instructional sequence were tested at the end of instruction, and then at a later date were brought back to be tested again on the same material, a single linear transformation would be expected to equate each individual's scores on the two tests if the same unidimensional trait space was in operation at both times of testing. This would occur because the metric underlying the trait space would have retained its interval properties (Lord & Novick, 1968) with the passage of time and no further instruction would have occurred that might change the ordering of the individuals in terms of achievement level.

Further, if the same group were brought back for additional tests at later dates, a linear trend should be found in the comparison of any two testing periods, provided that the latent space did not change. If the latent space did vary, a linear relation between the two measures of achievement level would not be expected. Thus, if a linear relationship is not observed between scores obtained at and after the peak of instruction, the conclusion can be drawn that two different traits were being evaluated at the two testing times.

This study was concerned with determining whether the unidimensional space defining achievement at the peak of instruction was sufficient to describe achievement after the peak of instruction.

Method

Testing Procedure

Testing at the peak of instruction was a requirement for students enrolled in the same undergraduate survey course in biology as in Study 1. The peak-of-instruction test data were from the required first midquarter examinations administered in a 2-day period to all students enrolled in the course in the fall academic quarter of 1976 and in the winter quarter of 1977.

Testing after the peak of instruction was implemented by the Computerized Adaptive Testing Project using volunteers from the same biology classes. These volunteers were given extra points toward their final course grade for participating in the research and were told that their level of performance on the computer-administered biology test would have no effect on their final grades. This testing began on the day following the midquarter examination and continued for approximately 1 month.

Peak-of-instruction testing (the required first midquarter examination) consisted of the conventional paper-and-pencil administration of 55 multiple-choice questions concerning the first three content areas in the course--"Chemistry," "The Cell," and "Energy." (For a more complete discussion of the course content and testing procedure, see Bejar, Weiss, & Kingsbury, 1977.)

The after-peak-of-instruction test (the second, voluntary test) was a computer-administered stradaptive test (Weiss, 1973) consisting of a maximum of 50 individually selected items chosen from the same item pool that was used to construct the required midquarter, including the same three content areas. (For a complete description of this testing procedure, see Bejar, Weiss, & Gialluca, 1977). The data used for this study were from 253 students from the fall quarter testing for whom achievement estimates from both tests and the date of the later test were available.

Since the adaptive and conventional tests were selected from the same content area pools, these two tests should have measured the same underlying dimension if the passage of time did not affect the latent space; and although differences in the precision of measurement between the two testing procedures were present (Bejar, Weiss, & Gialluca, 1977), they should not have affected the outcome of the present study.

Scoring

Peak-of-instruction achievement level estimates were obtained by scoring the students' midquarter item response data with the scoring program LINDSCO (Bejar & Weiss, 1979), which is designed to score conventional tests using item characteristic curve models. After-peak-of-instruction achievement estimates were obtained by scoring the stradaptive response vectors with the program ADADSCO (Bejar & Weiss, 1979), which is designed to score adaptive tests with item characteristic curve models. Since the maximum-likelihood logistic scoring method used in both programs is the same, the achievement level estimates obtained from the two programs are directly comparable.

Analysis

To determine whether the same latent space was operative after instruction that was operative at the peak of instruction, individuals' achievement levels at the peak of instruction were regressed on their achievement estimates after the peak of instruction. Since the later testing occurred over a period of a month, it was possible to analyze the effect of the passage of time on the relationship between achievement estimates. Since after-peak-of-instruction testing occurred only on weekdays, the weekends served as natural break points to divide the total group of students into four subgroups, each of which was tested during a different week in the month following the peak of instruction. Table 4 shows the total number of students tested each quarter, as well as the number tested in each week following the first midquarter examination.

If the time of testing after peak of instruction affects the latent space underlying testee responses, this effect may be studied by examination of the regression lines of peak testing on later testing using data from each of the 4 weeks of testing. To the extent that these regressions are parallel (i.e., exhibit no interaction between achievement level at the peak of instruction and the time of after-peak-of-instruction testing) and exhibit stable linear

Table 4
Total Number of Students Tested Each Quarter and Number Tested in Each Week Following Peak of Instruction (First Midquarter Examination)

Group	Quarter	
	Fall	Winter
Week 1	54	54
Week 2	83	90
Week 3	87	35
Week 4	29	6
Total	253	185

trends, it may be concluded that the latent space is stable and that the achievement metric was unchanged with the passage of time. As the time between testings lengthens, if increasing deviations from parallelism and/or linearity are observed, it may be concluded that the underlying trait space changed with the passage of time after instruction. Thus, both the parallelism and linearity of the regression of peak-of-instruction achievement level estimates on achievement level estimates obtained from after-peak-of-instruction testing were investigated.

Parallelism of regressions. For each of the 4 weeks of testing following the test administered at the peak of instruction, a separate regression line was obtained to predict individuals' later achievement levels from their peak-of-instruction achievement level estimates, using the subprogram REGRESSION contained in the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1970). In addition, the overall regression line was obtained, including all individuals regardless of the date of the later testing.

To statistically examine the parallelism of the regression lines from the 4 weeks following the classroom examination, it was necessary to determine whether the individual lines fit the data any better than the single overall regression line. This analysis used a test statistic described by Neter and Wasserman (1974). The statistic which determines whether the full model (F; the four individual regression lines) substantially reduced the sum of squares due to error (SSE) in the restricted model (R; the single overall regression line) in this application is

$$W = \frac{\mathrm{SSE}(R) - \mathrm{SSE}(F)}{6} \Big/ \frac{\mathrm{SSE}(F)}{N}, \qquad [1]$$

where N equals the number of individuals tested.

W is distributed as an F (6, N-8) distribution, and a significant value implies that the full model of four individual regressions is significantly more precise than the single restricted model. If the value of W is not significant, it implies that predictions of the students' achievement levels at the peak of instruction are just as good if the week of the later testing is ignored.

If the value of the statistic in Equation 1 is statistically significant, the individual regression lines are different in some respect from the overall regression line; it is then appropriate to test directly whether the individual regression lines differ significantly in slope. This may again be done through the use of Equation 1. In this instance, the restricted model becomes a single regression equation predicting individuals' later achievement estimates from their achievement estimates at the peak of instruction and from the week of the later testing. The full model uses these two predictors and adds the interaction of the two predictors to the model. A significant value for the statistic indicates a significant interaction between the predictor variables, indicating that the individual regression lines are not parallel. A significant result from this analysis would indicate a change in the latent space.

These analyses were implemented for both the fall quarter testing group and the winter quarter testing group in order to examine the stability of the results across independent groups.

Polynomial trend analysis. For each week of testing following the peak of instruction, it was desired to determine whether a linear trend existed and was sufficient to describe the prediction of the after-peak-of-instruction achievement estimate from the achievement estimate obtained at the peak of instruction. This was operationalized by fitting a fourth-degree polynomial regression equation to the data for each week of testing and separately determining the significance of each of the terms in the equations. To the extent that these regression equations exhibited an increasing trend toward curvilinearity as the time between testings increased, it could be inferred that the latent space was changing with time, causing a disruption in the interval properties of the original metric. If no such trend was observed, it could be concluded that the latent space remained stable as a function of time.

Regression equations were obtained from the REGRESSION subprogram in the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1970). Similar to the previous analysis, this analysis was implemented for both fall and winter quarters to permit replication of the results in independent groups.

Results

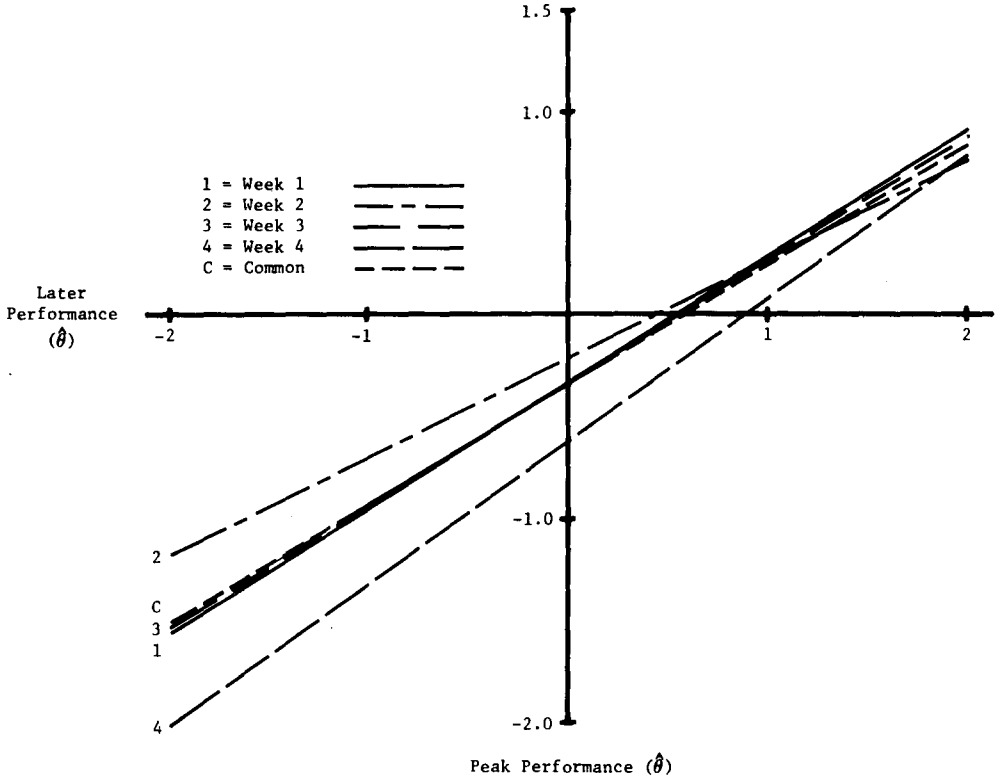
Parallelism of Regressions

Figure 5 shows individual regression lines obtained from each week of testing following the peak of instruction in the fall quarter (Figure 5a) and the winter quarter (Figure 5b), as well as the restricted regression line across weeks. Table 5 shows the sum of squares due to error and other descriptive statistics for each of the regression lines shown in Figure 5.

Using these sums of squares, the test for coincidence of regression in Equation 1 resulted in an F-value of 1.48 for the fall quarter data. This value, with 6 and 245 degrees of freedom, had a probability of occurrence by random fluctuation of $.10<p<.25$. For the winter quarter data, the observed F-value was .90, with 6 and 177 degrees of freedom. The probability of obtaining an F-value at least this extreme by random fluctuation was $p>.25$.

Figure 5
Regression Lines for Each Week Following Peak of Instruction
(First Midquarter Examination)

(a) Fall Quarter



(b) Winter Quarter

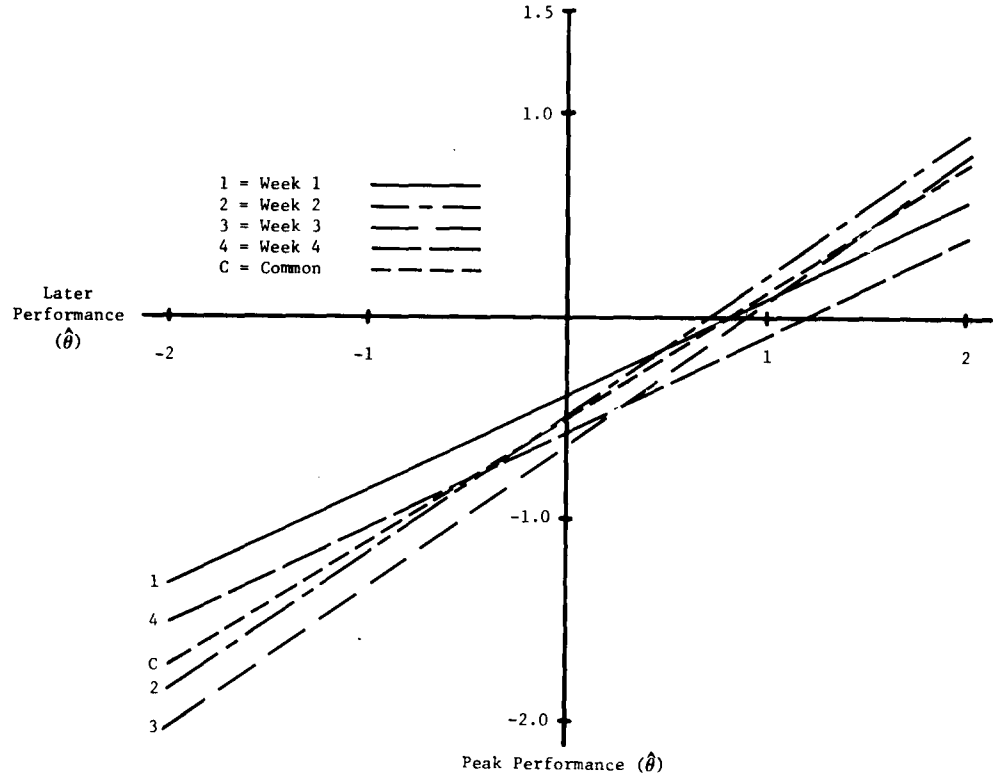


Table 5
Number of Subjects, Sum of Squares Due to Regression, and Sum of Squares Due to Error for Week-by-Week Regressions and for Common Regressions, for Fall and Winter Quarters

Quarter and Week	*N*	Sum of Squares	
		Error	Regression
Fall Quarter			
Week 1	54	21.88	23.76
Week 2	83	33.50	23.52
Week 3	87	41.32	36.70
Week 4	29	12.54	15.11
Common	253	113.24	103.26
Winter Quarter			
Week 1	54	20.77	13.32
Week 2	90	58.34	39.94
Week 3	35	12.83	20.75
Week 4	6	1.10	.78
Common	185	95.88	76.28

Neither obtained *F*-value was large enough to justify rejecting the null hypothesis. This finding implies that the individual regression lines obtained by taking the time lapse after the peak of instruction into account predicted later scores no better than the single overall regression line. Since the lines did not differ significantly, it was not necessary to test for parallelism among individual regression lines.

Polynomial Trend Analysis

Table 6 shows the significance of each polynomial term for the separate regression equations from each week of testing following the peak of instruction, for fall and winter quarters. This table shows that during the fall quarter, the linear term in the regression equation was significant, with a value of $p<.001$ in every week of testing. In only one instance was any other term's contribution to prediction significant at any reasonable significance level (i.e., $p<.05$): The quadratic term in the regression equation calculated from the fourth week of testing showed a significant contribution.

For the winter quarter the linear trend was again significant at the .001 level for the first 3 weeks of testing. The fourth week of testing resulted in the only nonsignificant linear trend in either quarter; this was probably due to the fact that the regression equation for the fourth week was based on only six students. It can also be seen that the quadratic term (the square of the peak achievement level estimate) in the regression equations obtained for each of the first 2 weeks of testing was a significant ($p<.05$) predictor of later performance for the winter quarter data. This trend was not evident in the third week of testing. Thus, a statistically significant quadratic trend was observed in the winter quarter data, but this trend did not increase as the time between testings increased. No other high order term contributed significantly to the prediction of later achievement level in either the fall or winter quarter data.

Table 6
Statistical Significance Level of Each Term of the Fourth Degree Polynomial Regression Equation Predicting Later Performance from Performance at the Peak of Instruction for Fall and Winter Quarters

Quarter and Week	*N*	Linear	Quadratic	Cubic	Quartic
Fall Quarter					
Week 1	54	$p<.001$	.358	.927	.693
Week 2	83	$p<.001$	.588	.326	.134
Week 3	87	$p<.001$	.510	.525	.518
Week 4	29	$p<.001$	.042	.748	.402
Winter Quarter					
Week 1	54	$p<.001$	.010	.671	.600
Week 2	90	$p<.001$	.018	.833	.589
Week 3	35	$p<.001$	.260	.541	.207
Week 4	6	.167	*a*	*a*	*a*

*a*Results not reported due to small *N*.

Discussion

Since the regression lines obtained in the different weeks of after-peak-of-instruction testing did not differ significantly from one another, the results of these analyses did not support the hypothesis that the achievement metric changed as a function of the time lapse between the peak of instruction and later achievement testing. The data indicate that the trait space was stable to the limit of the power of this analysis. It can also be seen from the data (see Figure 5) that student achievement measured after the peak of instruction was shifted in a linear manner from that measured at the peak of instruction; this is indicated by the nonzero intercepts and nonunit slopes of the overall regression lines from both quarters. The difference may be due to a lack of motivation or preparation for the after-peak-of-instruction tests, since that testing was voluntary and the scores on that test had no effect on students' course grades. In both quarters, however, there was no evidence of any change in the latent space with the passage of time.

The polynomial trend analysis indicated that in each week of testing following the peak of instruction, the maximum-likelihood estimates of achievement level at the peak of instruction were a significant linear predictor of later achievement levels. This finding was consistent across academic quarters. For the fall quarter the quadratic trend was significant only in the final week of testing. For the winter quarter the quadratic trend was a significant predictor in the first 2 weeks of testing, but not in the third week. These inconsistent findings imply that the quadratic trend observed may be a sample artifact.

The results from this analysis indicate that the only polynomial trend that acted as a consistent indicant of achievement was the linear term. It is, therefore, probable that the metric underlying individual testee response had not changed in any increasingly nonlinear manner, as would be expected if

the time between testings systematically affected the metric along which achievement was being measured, since no such systematic trend was noted.

Conclusions

The consistent findings of the analyses of the effect of time of after-peak-of-instruction testing on the measurement of achievement are as follows:

1. A single linear regression, using only prior performance as a predictor, was as efficient for the prediction of later performance as were four regression lines that took into account the time elapsed between peak of instruction and later testing.

2. The linear prediction trend was consistently significant in each week of testing following the peak of instruction.

3. No nonlinear prediction trend was consistently significant across all weeks of testing.

4. No significant increase in the significance of nonlinear prediction trends was observed with the increase of time elapsed between peak of instruction and later testing.

These findings lead to the conclusion that there was no evidence to support the hypothesis that the achievement variable changed as a function of the time lapse between peak of instruction and measurement of an individual's achievement level. The unidimensional ICC-based variable that had been used to measure achievement of individuals at the peak of instruction seemed to adequately describe the achievement of individuals as much as a month after the peak of instruction.

IMPLICATIONS AND LIMITATIONS

Implications

The results of these studies have implications for the measurement of achievement using both the pretest-test and test-posttest paradigms. The data suggest that there may be metric problems in the application of ICC item paramaters based on peak-of-instruction data to pretest data. Both the item difficulty and discrimination parameters estimated at the pretest (prior to instruction) differed substantially from those estimated at the peak of instruction. This result was reinforced by factor analyses of item sets obtained prior to instruction and at the peak of instruction; the variance accounted for by the first factor was considerably less at the pretest than it was at the peak of instruction. The implication of these results is that ICC-based pretest and peak-of-instruction achievement measurements may not be on the same dimension. Thus, the achievement variable measured at the pretest may be a different variable than that measured at the peak of instruction.

The importance of this finding, if it can be replicated in other data sets, is to call into question the utility of the pretest-test model for measuring gains in achievement. If the pretest achievement variable is, in fact, a different variable from that measured at the conclusion of instruction, it

is inappropriate to compute individual or group gain or change scores as indicants of growth in achievement. Such change scores would be completely useless in providing reliable estimates of growth in achievement levels because of the differences in the variables involved at the two points in time.

Contrary to the negative implications of these data for measuring achievement in a pretest-test paradigm, results of the second study support the use of an ICC-based test-posttest paradigm for the measurement of retention, at least within the 1-month time interval studied. Data from the second study showed that ICC-based achievement level estimates taken as much as a month after the peak of instruction were consistently linearly related to achievement level estimates taken at the peak of instruction. Thus, the data indicate that these posttest measurements were on the same ICC metric as the peak-of-instruction achievement level estimates. The data did show a level difference in the achievement estimates, but this might have resulted from design aspects of the study. Should future studies replicate this result (with or without the level difference), the data imply that gain (or loss) scores measuring retention after the peak of instruction may be meaningfully determined using ICC-based approaches.

The positive findings from the after-peak-of-instruction data also are in support of the potential of computerized adaptive testing for applications in the measurement of achievement. The data indicated linear relationships among ICC-based achievement level estimates obtained up to 4 weeks after instruction. Thus, even with limited availability of testing terminals, which might be characteristic of adaptive testing in certain instructional environments, it may be possible to obtain equivalent achievement estimates for students tested as long as a few weeks after the material was covered in a course. This should minimize the cost of an adaptive testing system and make its use economically feasible for classrooms of all sizes. Further research will be necessary, of course, to determine whether the observed mean differences in student performance after the peak of instruction were due to the motivational factors characteristic of voluntary participation.

Limitations

Both of the studies reported above were done within the context of a single undergraduate survey biology course. This limits the generalizability of the studies in several ways. For example, the weak factor structure noted in students' responses prior to instruction may be due to the fact that this was an introductory course. It is very possible that a more advanced course might show a strong prior-to-peak-of-instruction factor structure which is similar to the peak-of-instruction factor structure. In addition, different items drawn from the same content pool were used in the prior-to-instruction and peak-of-instruction factor comparisons. Future studies should compare the factor structures of the same items prior to instruction and at the peak of instruction.

Further limitations in the constancy of the testing procedures might have added some biases to the conclusions drawn. The pretest measure was required of all students attending the first lecture of the class, but it did not affect students' grades in the course. The first midquarter examination (peak-of-in-instruction test) was required of all students in the course and did have a bearing on the students' grade in the course. The after-peak-of-instruction measure was a voluntary test which allowed students to add extra credit points

to their course grade. In addition, this test was a computer-administered stradaptive test, whereas the first two tests were administered in conventional paper-and-pencil format. These differences, both methodological and motivational, may have added some unknown amount of bias to the results of the studies. Thus, replication of these studies in other instructional environments, and with revisions in the research design, is appropriate.

REFERENCES

Bejar, I. I., & Weiss, D. J. A construct validation of adaptive achievement testing (Research Report 78-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1978.

Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1979. (NTIS No. AD A067928).

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A047495).

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828).

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A046062).

Cronbach, L. J., & Furby, L. How we should measure "change"--or should we? Psychological Bulletin, 1970, 74,68-80.

Harris, C. W. (Ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Kingsbury, G. G., & Weiss, D. J. Relationship among achievement level estimates from three item characteristic curve scoring methods (Research Report 79-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1979.

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

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (NTIS No. AD A022964).

Neter, J., & Wasserman, W. Applied linear statistical models. Homewood, IL: Richard D. Irwin, 1974.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences. New York: McGraw-Hill, 1970.

Urry, U. W. A five-year quest: Is computerized adaptive testing feasible? In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Personnel Research and Development Center, PS-75-6). Washington, DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-00940-9).

Vale, C. D. Problem: Strategies of branching through an item pool. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975. (NTIS No. AD A018675).

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1975. (NTIS No. AD A020961).

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376).

APPENDIX: SUPPLEMENTARY TABLES

Table A
Parameter Estimates for Items Used in
Study of Peak/Peak Parameter Correspondence

Item Number	First Administration				Second Administration			
		Parameter				Parameter		
	Test*	*a*	*b*	*c*	Test*	*a*	*b*	*c*
3002	WF	.82	.13	.14	SF	.87	.12	.27
3034	W1	1.01	.37	.28	S1	.85	-.29	.13
3038	W1	1.58	-.56	.28	S1	1.20	-1.06	.16
3201	W1	1.07	-1.34	.23	S1	.85	-1.74	.18
3206	W1	.74	1.51	.21	S1	.75	1.57	.32
3216	W1	1.27	-.62	.18	S1	1.17	-.60	.15
3218	W1	.82	.58	.12	S1	.80	.34	.14
3237	WF	1.54	-.37	.18	SF	1.58	-.11	.43
3241	W1	1.12	2.48	.24	S1	.91	2.09	.17
3414	W1	.88	2.29	.32	S1	1.40	1.96	.30
3651	W2	.81	2.27	.44	S2	.95	2.31	.52
3812	W2	.74	-.66	.11	S2	.82	-.63	.13
3909	W2	1.34	.77	.38	S2	.90	1.12	.36
4006	WF	.84	-.59	.16	SF	1.05	-.19	.27
4036	WF	1.24	-.61	.23	SF	.95	-1.30	.17
4044	WF	.80	-.12	.38	SF	.80	-.60	.13
4229	WF	1.36	-.45	.38	SF	1.64	-.92	.17
4238	WF	.83	1.54	.42	SF	.83	1.47	.43

*W=Winter Quarter 1976; S=Spring Quarter 1976; 1=First Midquarter Examination; 2=Second Midquarter Examination; F=Final Examination

Table B

Interitem Correlations among 21 Items Selected from Pretest (Lower Triangle) and First Midquarter Examination (Upper Triangle)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	--	23	32	21	28	03	18	06	20	20	39	39	18	40	34	27	08	26	32	16	22
2	11	--	25	14	13	16	16	12	15	20	16	09	15	26	22	27	08	16	25	11	03
3	-11	02	--	15	28	22	22	16	23	16	22	24	06	15	29	31	12	12	29	12	11
4	04	-02	13	--	15	14	13	06	08	19	14	23	-02	20	14	24	19	12	18	15	11
5	11	00	01	05	--	55	25	34	36	37	39	21	21	26	38	25	23	12	39	13	21
6	01	-03	05	07	05	--	18	21	28	21	23	19	19	21	25	27	20	22	39	-04	10
7	02	06	03	10	00	10	--	11	23	00	19	15	15	13	13	19	10	06	29	13	12
8	07	-07	09	12	05	12	17	--	33	22	26	22	15	16	20	28	16	11	30	07	15
9	00	01	-05	-04	12	01	-02	04	--	27	30	19	24	08	21	28	19	19	31	14	17
10	07	02	03	03	18	05	04	01	08	--	26	15	27	35	23	34	14	15	38	17	10
11	19	01	01	06	12	-03	08	23	08	09	--	26	19	33	35	33	14	27	40	09	20
12	11	-03	03	12	05	00	08	11	-02	04	16	--	14	13	18	28	23	22	22	12	10
13	17	06	15	57	14	13	20	24	-01	04	22	14	--	26	24	15	-01	20	30	02	06
14	-04	03	10	05	-02	04	01	02	-05	06	03	-03	11	--	22	16	06	12	42	-01	11
15	24	-02	03	19	13	01	15	23	02	-02	22	09	24	09	--	36	13	22	38	-05	22
16	06	00	10	08	04	00	11	12	04	12	08	14	11	03	03	--	19	28	45	19	21
17	-01	00	01	04	-10	14	01	04	-10	-01	03	02	06	-01	-03	-13	--	16	19	19	16
18	20	01	08	21	06	00	19	14	-01	07	06	17	21	05	11	17	08	--	29	12	05
19	03	05	01	04	07	08	18	05	-09	-06	05	03	00	01	04	10	02	09	--	03	24
20	03	07	04	04	09	-02	01	10	00	07	05	10	11	16	07	-01	06	08	-02	--	15
21	19	03	02	09	07	06	16	03	06	-04	10	07	15	-03	11	-05	06	19	-01	02	--

Note. Item numbers are arbitrary; different items were selected from the same content areas at both Pretest and First Midquarter.

Table C
Five-Factor Solutions Prior to Instruction (Pretest)
and at Peak of Instruction (First Midquarter Examination)

Variable	Factor 1	2	3	4	5
Prior to Instruction					
1	.32	.38	-.26	.00	-.18
2	.04	.05	-.06	.01	-.13
9	.16	-.18	.18	-.01	.03
11	.54	-.40	-.03	-.11	-.12
15	.21	.19	-.01	-.22	-.01
17	.14	-.09	-.00	.13	.09
18	.32	.03	.10	.29	-.01
19	.41	.08	.14	.11	.41
22	.02	.17	.01	-.19	.04
23	.12	.12	.12	-.18	-.05
24	.36	.26	-.04	-.09	.19
25	.27	.11	.08	.04	-.03
27	.75	-.32	-.11	-.18	-.03
29	.11	-.10	.07	-.05	.06
30	.41	.15	-.12	-.03	.15
32	.25	.15	.51	-.04	-.17
33	.06	-.14	-.15	.24	.08
36	.41	.08	.07	.20	-.22
37	.12	.05	.12	.24	-.04
38	.16	.01	-.01	-.06	.07
39	.25	.11	-.24	.15	-.11
Peak of Instruction					
1	.59	.68	-.08	-.33	.01
4	.36	.11	-.04	.16	.00
7	.44	.07	.10	-.03	-.15
10	.31	.12	.15	.03	.07
13	.70	-.46	-.04	-.48	.15
14	.49	-.38	-.06	-.03	-.14
16	.33	-.01	.09	-.07	-.11
19	.41	-.20	.07	.07	-.01
22	.49	-.16	.13	.03	-.02
25	.51	-.07	-.10	.23	.40
28	.58	.07	-.05	-.04	-.04
31	.43	.15	.19	-.05	-.10
34	.36	-.04	-.22	.12	.04
37	.48	.16	-.39	.05	.20
40	.54	.01	-.13	-.00	-.19
41	.60	.04	.20	.28	-.11
43	.31	-.08	.30	.00	.02
46	.38	.10	.06	.14	-.11
49	.69	-.05	-.16	.19	-.08
52	.20	.12	.46	.01	.32
55	.31	.03	.12	-.07	-.01

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

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

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 CDR Robert S. Kennedy
Naval Aerospace Medical and Research Lab
Box 29407
New Orleans, LA 70189

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION & TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs (Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc (OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

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

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 LCOL Gary Bloedorn
Training Effectiveness Analysis Division
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, NM 88002

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. 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 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

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

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

Marines

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

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM
Ft. 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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

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

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Warner Birice
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Educational Psychology
University of Illinois
Champaign, IL 61820

1 Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairmar
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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

1 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

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

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

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

1 Dr. Robert Tsutakawa
Dept. of Statistics
University of MIssouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979. (NTIS No. AD A068176)
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979. (NTIS No. AD A067928)
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979. (NTIS No. AD A067752)
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978. (NTIS No. AD A064352)
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978. (NTIS No. AD A062797)
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978. (NTIS No. AD A059436)
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977. (NTIS No. AD A047495)
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977. (NTIS No. AD A046062)
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977. (NTIS No. AD A044828)
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977. (NTIS No. AD A041090)
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977. (NTIS No. AD A041084)
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977. (NTIS No. AD A038114)
Final Report: Computerized Ability Testing, 1972-1975. April 1976. (NTIS No. AD A024516)
76-5. Effects of Item Characteristics on Test Fairness. December 1976. (NTIS No. AD A035393)
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976. (NTIS No. AD A027170)
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976. (NTIS No. AD A028147)
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976. (NTIS No. AD A024422)
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (NTIS No. AD A022964)
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975. (NTIS No. AD A020961)
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (NTIS No. AD A018675)
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975. (NTIS No. AD A018758)
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (NTIS No. AD A013185)
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975. (NTIS No. AD A007572)
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (NTIS No. AD A006733).
74-5. Strategies of Adaptive Ability Measurement. December 1974. (NTIS No. AD A004270)
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974. (NTIS No. AD A001230)
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (NTIS No. AD 783553)
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (NTIS No. AD 781894)
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974. (NTIS No. AD 773961)
73-3. The Stratified Adaptive Computerized Ability Test. September 1973. (NTIS No. AD 768376)
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973. (NTIS No. AD 757788)

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.
Copies of these reports are available, while supplies last, from:
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

AN ADAPTIVE TESTING STRATEGY FOR MASTERY DECISIONS

G. Gage Kingsbury

and

David J. Weiss

RESEARCH REPORT 79-5
SEPTEMBER 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Army Research Institute, Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, and the Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* An Adaptive Testing Strategy for Mastery Decisions		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* G. Gage Kingsbury and David J. Weiss		8. CONTRACT OR GRANT NUMBER*(s)* N00014-76-C-0627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.: NR150-389
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE September 1979 13. NUMBER OF PAGES 36
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 in 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

This research was supported by funds from the Army Research Institute, Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, and the Office of Naval Research, and monitored by the Office of Naval Research.

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

latent trait test theory	achievement testing	testing
item response theory	computerized testing	tailored testing
response-contingent testing	adaptive testing	programmed testing
individualized testing	sequential testing	automated testing
	branched testing	

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

In an attempt to increase the efficiency of mastery testing while maintaining a high level of confidence for each mastery decision, the theory and technology of item characteristic curve (ICC) response theory (Lord & Novick, 1968) and adaptive testing were applied to the problem of judging individuals' competencies against a prespecified mastery level to determine whether each individual is a "master" or a "nonmaster" of a specified content domain. Items from two conventionally administered classroom mastery tests administered in a

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

military training environment were calibrated using the unidimensional three-parameter logistic ICC model. Then, using response data originally obtained from the conventional administration of the tests, a computerized adaptive mastery testing (AMT) strategy was applied in a real-data simulation.

The AMT procedure used ICC theory to transform the arbitrary "proportion correct" mastery level used in traditional mastery testing to the ICC achievement metric in order to allow the adaptation of the test to each trainee's achievement level estimate, which was calculated after each item response. Adaptive testing continued until the 95% Bayesian confidence interval around the trainee's achievement level estimate failed to contain the prespecified mastery level. At that point testing was terminated, and a mastery decision was made for the trainee.

Results obtained from the AMT procedure were compared to results obtained from the traditional mastery testing paradigm in terms of the reduction in mean test length, information characteristics, and the correspondence between decisions made by the two procedures for three different mastery levels and for each of the two tests. The AMT procedure reduced the average test length 30% to 81% over all circumstances examined (with modal test length reductions of up to 92%), while reaching the same decision as the conventional procedure for 96% of the trainees.

Additional advantages and possible applications of AMT procedures in certain classroom situations are noted and discussed, and further research questions are suggested.

CONTENTS

Acknowledgments

Test data utilized in this study were obtained from Air Force personnel enrolled in the Weapon Mechanics course at the Lowry Air Force Base Technical Training Center from 1977 to 1978. The authors extend their appreciation to Brian Waters and Larry Click of the Air Force for making these data available and to Joel Brown for arranging for the data to be transferred in a usable form.

Technical Editor: Barbara Leslie Camm

An Adaptive Testing Strategy for Mastery Decisions

During the past 15 years, considerable interest in the psychological and educational measurement community has been directed toward the evaluation of student competency in various fields of study. In the simplest case, competency in a field has been operationalized as some minimum skill level above which a student is declared a "master" and below which a student is declared a "nonmaster." Mastery testing has been developed as an implementation of the more general criterion-referenced test interpretation model formulated by Glaser and Klaus (1962) and expanded upon by many since then (e.g., Hambleton, Swaminathan, Algina, & Coulson, 1978; Popham, 1971; Popham & Husek, 1969).

"Mastery" has typically been defined by subject matter experts as the minimum percentage of items that a student should be able to answer from a given set of test items in order to be classified as proficient. Therefore, a student who correctly answered only the minimum acceptable percentage of items on a test of this type would be declared a master, and a student who correctly answered one item less would be declared a nonmaster in the subject matter area. So that all of the mastery decisions made would be comparable, mastery testing has traditionally required all students to answer the same set of test questions.

This approach to mastery testing has several problems. First, a student whose test score is far above the specified cutoff score would be said to be a master of the subject matter; similarly, a student whose score was just barely above the cutoff score would also be declared a master, but presumably that decision would be made with less confidence. Thus, classical mastery testing results in different levels of intuitive confidence for students whose raw scores fall at different distances above or below the cutoff, which results in decisions with different dependabilities for students with different raw scores.

This problem has been discussed on the group level by Livingston (1972) in a study discussing the reliability of criterion-referenced tests as a function of the mean score level of the testee group. Hambleton and Novick (1973) and Davis and Diamond (1974) have specified methods to develop cutoff rules designed to yield certain desired ratios of false positive and false negative decisions through the use of the differential accuracy of decisions made at different raw score levels, but little research has been directed toward equalizing the confidence levels in decisions made by a mastery test across all levels of performance. Hambleton and Novick (1973) have suggested that the use of Bayesian point estimation of students' mastery scores might improve the accuracy of mastery decisions; it will be shown in this report that the use of Bayesian confidence interval estimates may be useful in equalizing the confidence in decisions made across all levels of observed performance.

A second problem with the classical mastery testing paradigm is that each student tested is given the same set of test questions, even though the set of questions may be inappropriate for any reasonably precise measurement at some

achievement levels. In the mastery testing area, attempts have been made to adapt the test to each student (e.g., Ferguson, 1970); but these attempts have almost universally assumed that all items administered were of equal quality. It is possible, through the use of item characteristic curve (ICC) response theory (Lord & Novick, 1968), to distinguish between items which yield different amounts of information concerning different trait levels.

Several authors (e.g., Bejar, Weiss, & Gialluca, 1977; McBride & Weiss, 1976; Urry, 1977) have demonstrated that adaptive testing procedures using ICC response theory can reduce test length with no reduction in measurement precision. These testing procedures adapt the difficulty and information characteristics of each individual's test by drawing from large item pools items that are matched to the individual's estimated trait level. These results indicate that by making use of all of the information available about the test items and the individual's estimated achievement levels, the application of adaptive testing procedures using ICC response theory to a traditional mastery testing situation might result in a decrease in the test length needed to make confident decisions concerning each individual's mastery status.

Objectives

This report describes the design and application of an adaptive mastery testing strategy that eliminates these problems of the traditional mastery testing approach. The adaptive mastery testing strategy is designed to reduce the average test length for each student, while equalizing the level of confidence in decisions made across the entire range of the achievement continuum. This report compares the performance of the conventional and adaptive mastery testing procedures within the context of one course of instruction in terms of efficiency, information characteristics, and level of correspondence between masery decisions.

The Adaptive Mastery Testing Procedure

The adaptive mastery testing (AMT) procedure is designed to administer achievement test items selected from a classical mastery test, but not all items are administered to each student. The test items administered to a given student are selected to provide the most information concerning the achievement level of that student. Mastery decisions are made with a specified degree of confidence for each student, using a cutoff point prespecified on the achievement continuum.

There are three important components of the AMT procedure. The first involves converting the mastery level to the achievement metric. The second component is the item-selection technique used to determine which items should be administered to a specific student. The final component of the AMT strategy involves the manner in which the mastery decision is made and the degree of confidence that can be placed in the decision once it has been made.

Mastery and the Achievement Metric

The classical mastery testing procedure specifies a percentage of the items on a test that must be correctly answered by a student in order to be declared a master. Using ICC theory, it is possible to generate an analogue to the "percentage" cutoff of classical theory for use in adaptive testing. This is nec-

essary, since in an adaptive test each individual will tend to answer about 50% of the items correctly, given a large enough item pool, because the items administered will be selected to be close to the individual's achievement level (Vale & Weiss, 1975; Weiss, 1973). The ICC analogue of proportion correct is based on the use of the test characteristic curve (TCC). The TCC is the function that relates the ICC achievement continuum to the expected proportion of correct answers that an individual at any achievement level may be expected to obtain if all of the items on the test were administered.

For this study the assumption was made that a three-parameter logistic ogive would describe the functional relationship between the latent trait (achievement) and the probability of observing a correct response to any of the items on the test. This assumption yields a TCC of the following form:

$$E(P|\theta) = \sum_{i=1}^{n} \left[c_i + (1 - c_i) \left(\frac{\exp[1.7a_i(b_i-\theta)]}{\exp[1.7a_i(b_i - \theta)]+1} \right) \right] \Big/ n \qquad [1]$$

where

$E(P|\theta)$ = the expected value of the proportion of correct answers observed on the test, given an achievement level;
a_i = the estimate of the ICC discrimination parameter for item i;
b_i = the estimate of the ICC difficulty parameter for item i;
c_i = the estimate of the lower asymptote of the ICC for item i;
n = the number of items on the test; and
θ = a given achievement level.

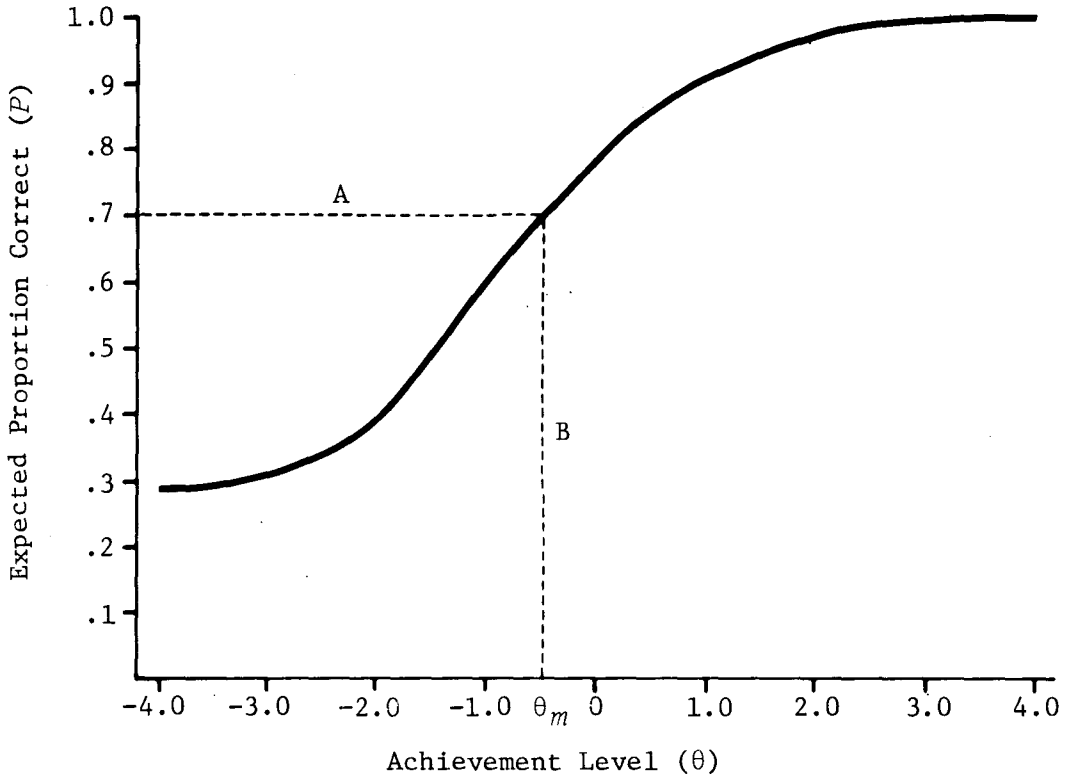
Thus, as Equation 1 indicates, the expected proportion correct at a given level of achievement (θ) is the average, over all items in the test, of the probability of a correct response for each item, given the three ICC item parameters for each item and assuming a logistic ICC.

This monotonically increasing function permits relating any achievement level to its most likely proportion correct or, more importantly in this context, determining the achievement level (θ) which will most probably result in any given proportion of correct answers. An example of the use of the TCC in determining an achievement level that is comparable to a desired "percentage" cutoff is shown in Figure 1 using a hypothetical TCC. To determine a level of achievement that corresponds to, for example, a 70% mastery level on the test items which comprise the TCC, these steps would be followed:

1. Draw a horizontal line (line A in Figure 1) from the P=.7 mark on the vertical (expected proportion correct, or P) axis of the TCC plot to the TCC.

2. Drop a vertical line (line B) from the point of intersection of the TCC and the horizontal line drawn in Step 1 to the horizontal (achievement level, or θ) axis. This point (θ_m) on the achievement level axis is designated the mastery level using the achievement metric.

Figure 1
Hypothetical Test Characteristic Curve Illustrating
Conversion from a Proportion Correct Mastery Level
to the Achievement Metric



3. The cutoff point specified in Step 2 may now be used to make mastery decisions in place of the P=.7 mastery level originally specified. Once the mastery level is expressed in the achievement metric (θ), rather than in terms of proportion correct, it is no longer necessary to administer all the items in the test to obtain an achievement level estimate for an individual--and a corresponding mastery decision. An achievement level estimate can then be obtained using any subset of items from the original test, provided that the individual's item responses are scored with a method that will put the achievement level estimate on the same metric as the TCC. Any ICC-based scoring procedure (Bejar & Weiss, 1979), in conjunction with the original item parameter estimates, will result in an achievement level estimate which will be on the θ metric.

This procedure allows conversion of any desired proportion correct mastery level to the θ metric. Once this transfer is made, ICC theory and adaptive testing strategies may be used to increase the efficiency of mastery testing techniques.

Adaptive Item Selection and Scoring

To make mastery testing a more efficient process, the objectives of the AMT strategy were (1) to reduce the length of each student's test by elimi-

nating test items which provided little information concerning the student's achievement level and (2) to terminate the AMT procedure after enough information had been obtained so that the mastery decision could be made with a high degree of confidence.

To operationalize the first objective, items were selected to be administered to student at each point during the testing procedure on the basis of the amount of information that the item provided concerning the student's achievement level estimate at that point in testing. The administration of the test item which provides the most information concerning the student's present achievement level estimate should provide the most efficient use of testing time. A procedure that selects and administers the most informative item at each point in an adaptive testing procedure was described by Brown and Weiss (1977), and this procedure was used in the present study. This procedure uses an adaptive maximum information search and selection (MISS) technique for the sequential selection of test items to be administered to each individual.

Item selection. The information that an item provides at each point along the achievement continuum can be determined from the ICC parameters of the item. Using the unidimensional three-parameter logistic ICC model (Birnbaum, 1968) to describe responses to the five-alternative multiple-choice items used in this study, the information available in any item is (Birnbaum, 1968, Equation 20.4.16)

$$I_i(\theta) = (1-c_i)D^2a_i^2\psi^2\,[DL_i(\theta)] \Big/ \{\psi[DL_i(\theta)] + c_i\Psi^2[-DL_i(\theta)]\} \qquad [2]$$

where

$I_i(\theta)$ = the information available from item i at any achievement level θ;

a_i = the ICC discrimination parameter of the item;

c_i = the lower asymptote of the ICC for the item;

D = 1.7, a scaling factor used to allow the logistic ICC to closely approximate a normal ogive;

$L_i(\theta) = a_i(\theta - b_i)$, where b_i is the ICC difficulty parameter of the item;

ψ = the logistic probability density function; and

Ψ = the cumulative logistic function.

If it assumed that the achievement level estimate ($\hat{\theta}$) is the best estimate of the true achievement level (θ), item information levels of each of the items not yet administered can be evaluated using $\hat{\theta}$ at any point during the test. The item which has the highest information value at the individual's current level of $\hat{\theta}$ is thus chosen to be administered next. Appendix A (adapted from Brown & Weiss, 1977) gives an example of the use of the MISS procedure to select items.

Estimation of θ. For this study a Bayesian estimator (Owen, 1969) of the student's achievement level ($\hat{\theta}$) was used. Details of the scoring procedure have been provided by Brown and Weiss (1977, pp. 4-5); Bejar and Weiss (1979) have provided an explanation and scoring programs for Owen's method.

Owen's θ estimation procedure has been shown to yield biased estimates of trait levels (Kingsbury & Weiss, 1979; Lord, 1976; McBride & Weiss, 1976). This bias may be attributed to the assumption of a normal distribution of θ in the population made by Owen's procedure (Lord, 1976) and/or to inappropriate prior information concerning θ on the individual level (Kingsbury & Weiss, 1979). The bias inherent in this scoring method may render the MISS technique less efficient than it would be under optimal conditions, and thereby may reduce the efficiency of the AMT technique as a whole.

To use MISS under optimal conditions, θ estimates should be obtained through the use of a maximum likelihood estimation technique, which yields asymptotically efficient estimates (Birnbaum, 1968). Maximum likelihood θ estimation techniques are not able, however, to obtain trait level estimates for consistent item response patterns (either all correct or all incorrect responses) or for item response patterns for which the likelihood function is extremely flat. Owen's Bayesian scoring method will yield an estimate for any response pattern. The inability of the maximum likelihood procedures to estimate θ for some response patterns mitigates against the use of a maximum likelihood estimation procedure in this situation, since it would be necessary to assign arbitrary θ estimates during the early stages of item selection and scoring. Thus, the Bayesian scoring procedure was used in order to obtain θ estimates for each student after each item administered by the adaptive testing procedure, even though some efficiency might have been lost in the AMT due to the bias inherent in the estimation procedure. Use of the Bayesian θ estimation procedure in this study also allowed the use of easily interpretable Bayesian confidence intervals to make the mastery decision.

Bayesian Confidence Intervals: Making the Mastery Decision

Any achievement level estimate ($\hat{\theta}$) obtained using ICC-based scoring of any subset of the items from the original test and their ICC item parameters will be on the same metric as the TCC for the original test. This allows immediate comparison between any achievement level estimate ($\hat{\theta}$) and any point on the achievement metric (e.g., θ_m). However, two different subsets of items may result in achievement level estimates that are not equally informative. For example, if one test consisted of many items that were too easy for a given individual and the other used the same number of equally discriminating items at about the appropriate difficulty level for that individual, the second test would yield a much more accurate achievement level estimate for that individual. Achievement level estimates that are on the same metric are comparable if their differential precision is taken into account. To do this, confidence interval estimates for the $\hat{\theta}$'s should be compared instead of the point estimates ($\hat{\theta}$). For this reason, the AMT strategy makes mastery decisions with the use of Bayesian confidence intervals.

After each item was selected using MISS and administered to a student, a point estimate of the student's achievement level ($\hat{\theta}$) was determined using Owen's Bayesian scoring algorithm and the responses obtained from all items previously administered. Given this point estimate and the corresponding variance estimate for the $\hat{\theta}$, also obtained using Owens' procedure (see Brown & Weiss, 1977, Equations 3 and 5, pp. 4-5), a Bayesian confidence interval may be defined such that:

$$\hat{\theta}_i - 1.96(\hat{\sigma}_i^2)^{\frac{1}{2}} \leq \theta \leq \hat{\theta}_i + 1.96(\hat{\sigma}_i^2)^{\frac{1}{2}}, \text{ with } P = .95, \qquad [3]$$

where $\hat{\theta}_i$ = the Bayesian point estimate of achievement level calculated following item i,

$\hat{\sigma}_i^2$ = the Bayesian posterior variance estimate following item i,

and θ = the true achievement level.

This statement may be interpreted as meaning that the probability that the true value of the achievement level parameter, θ, is within the bounds of the confidence interval is .95. Alternatively, it might also be concluded with 95% confidence that the true parameter value (θ) lies within the confidence interval. Confidence intervals at differing confidence levels can be constructed using appropriate z-values from a normal distribution in place of the 1.96 in Equation 3.

After this confidence interval has been generated, it can be determined whether or not θ_m, the achievement level earlier designated as the mastery level using the TCC (see Figure 1), falls outside the limits of the confidence interval. If it does not, another item is administered to the student, and the confidence interval is recalculated using the updated $\hat{\theta}$ and its updated variance. This procedure continues until, after some item has been administered, the confidence interval calculated does not include θ_m, the mastery level on the achievement continuum. At this point testing is terminated, and a mastery decision is made. If the lower limit of the confidence interval falls above the specified mastery level, θ_m, the student is declared a master. If, on the other hand, the upper limit of the confidence interval falls below θ_m, the student is declared a nonmaster. Given a finite size item pool, the testing procedure may, in some cases, exhaust the item pool before a decision can be made. This will occur for students with $\hat{\theta}$ values close to θ_m. It is possible to make a mastery decision for these students based simply on whether the Bayesian point estimate of their achievement level ($\hat{\theta}$) is above or below θ_m. However, for these students, mastery decisions will not be made with the same confidence levels as those made for students for whom the confidence interval falls completely above or below θ_m.

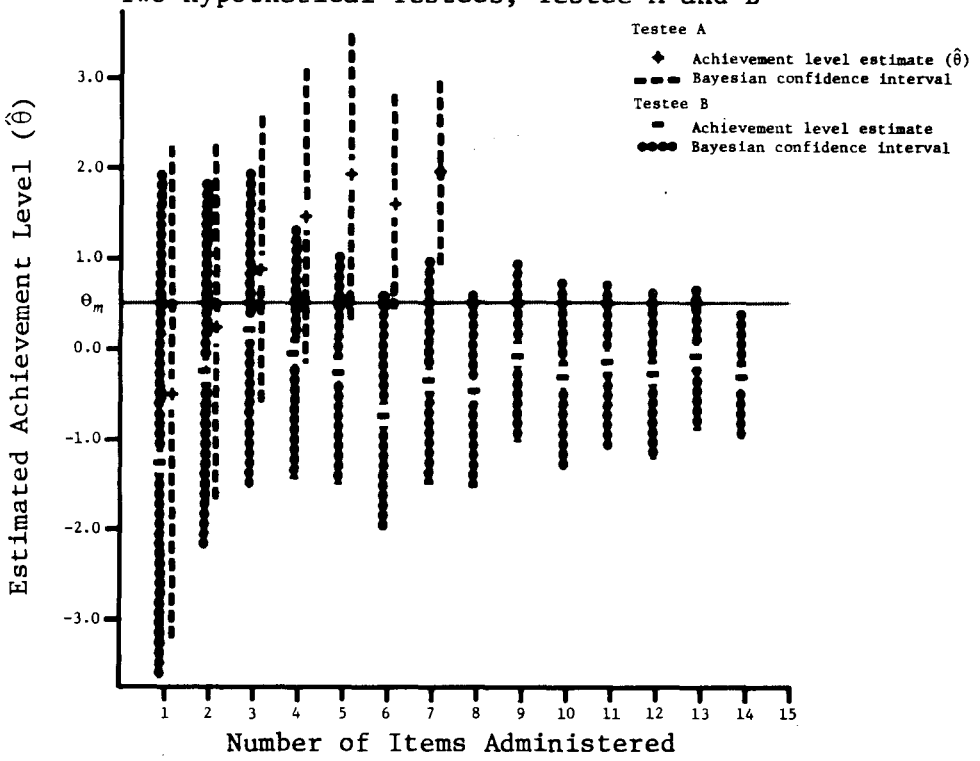
Illustration

Figure 2 shows the result of the AMT procedure for two hypothetical testees, A and B. Achievement level point estimates ($\hat{\theta}$) and error bands, which indicate the appropriate Bayesian confidence intervals, are shown for each testee after each item was administered. An arbitrary mastery level, $\theta_m = .50$, was chosen for this example; normally, however, the mastery level would be determined by the TCC transformation of an existent proportion correct mastery criterion.

For Testee A, the first θ estimate was below θ_m, but the confidence interval around this estimate contained θ_m. Thus, the θ estimate was not precise enough to make a confident decision; consequently, testing continued for Testee A. After each item was administered, a new θ estimate and a corresponding confidence interval were calculated. For the first 6 items administered to

Testee A, the confidence interval around the θ estimate contained θ_m, and testing continued. After the administration of the $7th$ item, the entire confidence interval around the θ estimate for Testee A was above θ_m. This implied that the θ estimate was precise enough to allow a confident decision to be made for Testee A. Testee A was declared a master at this point, and testing was terminated.

Figure 2
Example of the AMT Procedure: Achievement Level Point Estimates and Bayesian Confidence Intervals after Each Item Administered to Two Hypothetical Testees, Testee A and B



For Testee B, the same type of procedure was followed. For the first 13 items administered to Testee B, the confidence interval around $\hat{\theta}$ contained θ_m. The $14th$ item administered to Testee B resulted in a $\hat{\theta}$ and confidence interval which fell completely below θ_m. At that point, testing was terminated and Testee B was declared a nonmaster of the subject area.

It should be noticed that Testee A had a final θ estimate ($\hat{\theta}\simeq 1.9$) that was much closer to the mastery level than the final θ estimate for Testee B ($\hat{\theta}\simeq -.30$). Therefore, much more precise measurement was needed for Testee B than for Testee A to make mastery decisions with comparable confidence levels, and several more items were administered to Testee B than to Testee A, to obtain the additional precision needed in order to make the mastery decision.

Method

The AMT strategy was evaluated using real-data simulation (Weiss, 1973). In this approach, test item response data obtained from the administration of a conventional paper-and-pencil multiple-choice achievement test were used to simulate the administration of the AMT strategy. That is, items were selected

by the AMT strategy for each student from the conventional test already administered. Item responses obtained in the conventional test were used by the AMT strategy and scored as described above. If a mastery decision could not be made after a given item was used, another item from the conventional test was selected by the MISS approach, and the previously obtained item response was used by the AMT strategy. This procedure was continued until the AMT strategy could make a mastery decision or until all items in the conventional test pool had been administered.

Subjects and Tests

Item response data were obtained from trainees undergoing the Weapon Mechanics course at the Lowry Air Force Base Technical Training Center during 1977 and 1978. This course is computer-managed, and trainees proceed at their own pace through 13 well-specified blocks of instruction. During each block, several tests are given from which mastery decisions are made. Trainees are given several attempts to pass each test in each block.

For this study two block tests of different lengths were arbitrarily chosen to investigate the properties of the AMT procedure. Specifically, data used were the item responses of 200 trainees to the first test in the first block of instruction (Test 11) and the item responses of 200 trainees to the first test in the third block of instruction (Test 31). These tests consisted of 30 and 50 conventionally administered 5-alternative multiple-choice items, respectively. Only the trainees' performances in their first attempt to pass the tests were used for this study.

Fitting the ICC Response Model

Estimation of item parameters. The procedure used for the estimation of the three item parameters of the logistic ICC response model was developed by Urry (1976). This procedure obtains initial estimates for the discrimination (a) and the difficulty (b) parameters for an item through the use of a direct conversion of the classical item parameters and the individuals' raw scores (number correct). A value of the lower asymptote parameter (c) is found which minimizes a χ^2 goodness-of-fit statistic for the item. These initial values are made more precise through the use of an ancillary correction procedure (Fisher, 1950). To obtain more precise estimates of the parameters, the entire procedure is repeated replacing the individuals' raw scores with Bayesian modal estimates (Samejima, 1969) of their achievement levels.

Urry's item parameterization method excludes items which meet any of the following rejection criteria during the first stage of the procedure:

1. a less than .80,
2. b less than -4.00 or greater than 4.00, and
3. c greater than .30.

If an item is excluded on the basis of one of these criteria during the initial stage of the parameterization procedure, it receives no parameter estimates in either stage of the procedure. These restrictive criteria are removed after the first phase of the calibration, and no further culling of the items is done. Thus, the final values of the parameter estimates for those items which survive the first phase are not constrained by the rejection criteria.

Evaluating the fit of the model. To examine the usefulness and appropriateness of the unidimensional three-parameter logistic ICC model with data of the type provided by the Weapon Mechanics course, two questions were investigated:

1. Does factor analysis of the intercorrelations between item responses result in only a single common factor? That is, is the use of a unidimensional model justified by the presence of only a single nonrandom dimension?
2. Do parameter estimates obtained from these data correspond to the range of parameter estimates obtained in previous studies that have shown this type of model to be useful in increasing testing efficiency?

To answer the first question, principal axis factor analyses were performed separately on the data from Test 11 and Test 31. Matrices of item intercorrelations (phi coefficients) were calculated from the raw item-response data for the 200 trainees on each of the tests using the PEARSON CORR computer subroutine from the *Statistical Package for the Social Sciences* (*SPSS*; Nie, Hull, Jenkins, Steinbrenner, & Bent, 1970).

The resultant 30 × 30 (Test 11) and 50 × 50 (Test 31) item intercorrelation matrices were each factor analyzed by the iterative principal axis factor analysis subroutine from *SPSS*. The initial communality estimate for each of the items was the squared multiple correlation of the item with all other items in the test. The analysis iterated until successive communality estimates differed by a negligible amount.

To determine the amount of random variation in the final factor-analytic solutions, parallel analyses were conducted following the suggestion of Horn (1965). This entailed factor analyses of sets of random data that were generated to parallel the original data, using the same number of "items" and "subjects." Eigenvalues obtained for factors in the random data were used to determine whether factors obtained from the analysis of the real data were "true" factors or residual factors. If the eigenvalue of a factor obtained from the real data was larger than that for the corresponding random-data factor, the real-data factor was considered to be a true factor; but if the eigenvalue was similar to that obtained from the random-data factor, then the real-data factor was considered to be a residual factor of no real importance.

To answer the second question posed above, the parameter estimates obtained for these two tests were compared to the estimates obtained in two other studies (Bejar, Weiss, & Kingsbury, 1977; Brown & Weiss, 1977) that used a unidimensional three-parameter logistic ICC model to attempt to improve testing accuracy in achievement testing situations. Further comparisons were made between the parameter estimates obtained from the present data and the guidelines expressed by Urry (1977) to indicate whether the use of an adaptive testing item pool will improve the quality or efficiency of trait measurement. Urry's guidelines are as follows:

1. The a parameter estimates of the items in the pool should exceed .80.
2. The b parameter estimates should be widely and evenly distributed between -2.00 and +2.00.
3. The c parameter estimates should be less than .30.

To the extent that parameter estimates obtained from Tests 11 and 31 followed Urry's guidelines and showed close correspondence to other item pools that have

proven to be useful in adaptive testing, it could be concluded that the items used in this study would show some usefulness with the unidimensional three-parameter ICC model.

Simulation of AMT

In order to simulate the AMT strategy, a computer program was designed to "administer" the one item in the item pool (which included all of the items from the conventional test not rejected by the calibration procedure) providing the most information at a trainee's current level of $\hat{\theta}$. Each trainee began the test with $\hat{\theta}$ of 0.0 and a prior variance of 1.0. The trainee's response taken from his/her original responses to the conventional test was used by the Bayesian scoring routine to produce a new θ estimate. Then the item with the most information at this new $\hat{\theta}$ was chosen to be administered next. (No item was administered more than once to a trainee.) A new θ estimate was found using the trainee's response to this item, and then another item was chosen based on the new θ estimate.

The program continued to choose items to be administered until the trainee's $\hat{\theta}$ was shown to be either above or below a given mastery level, θ_m, with a prespecified degree of confidence. A 95% Bayesian symmetric confidence interval was calculated around the trainee's $\hat{\theta}$ after each item was administered. The AMT strategy continued until this confidence interval failed to include the prespecified mastery level; when this occurred, the AMT procedure was terminated. A lower limit of three items was set for the length of the AMT to avoid anomalous results that might occur from making mastery decisions based on a small number of item responses. For trainees for whom a mastery decision could not be made with the AMT procedure before all items were administered, mastery was determined by whether the final $\hat{\theta}$ was above or below θ_m.

During the simulation, three different mastery levels were used corresponding to proportion correct mastery levels of P=.7, .8, and .9. These mastery levels were calculated from the TCC for each test, as described above. To maximize the comparability between the conventional and adaptive mastery testing strategies, the conventional test was truncated to include only the items which were not rejected by the calibration procedure. In addition, the conventional test was scored by Owen's Bayesian scoring method, and the same mastery levels were used for both testing strategies.

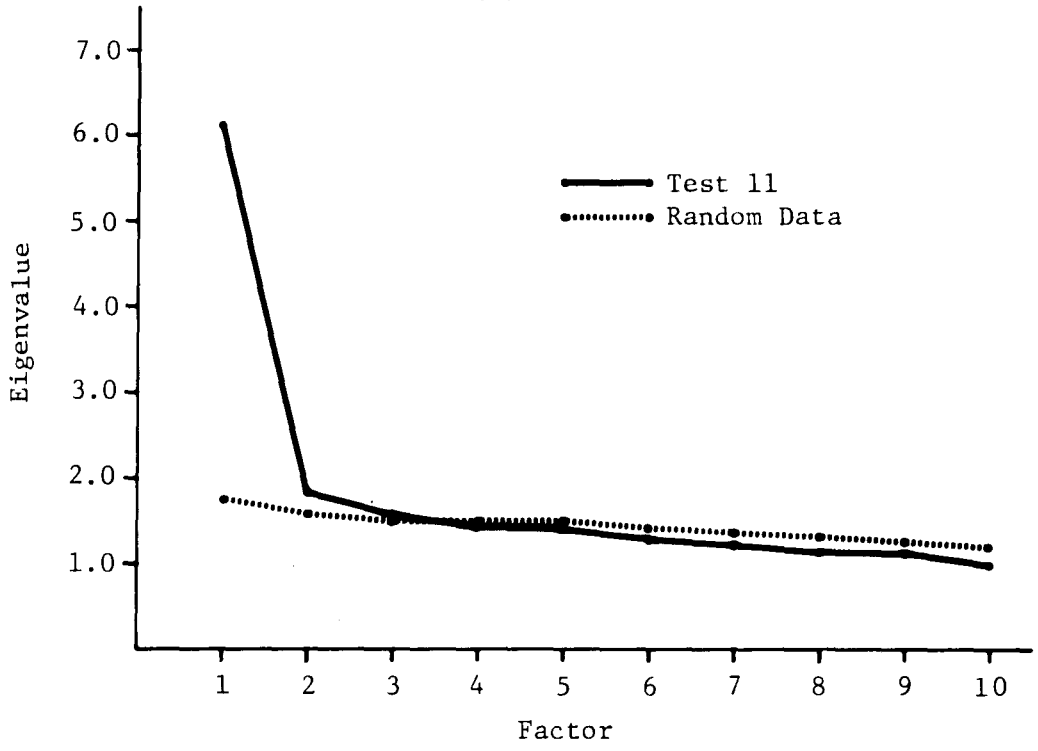
Comparison of Efficiency: AMT versus Conventional Testing

If the AMT strategy were a more efficient testing procedure than the conventional mastery testing procedure, it would reduce test length while administering items with high enough information to maintain a very high correlation between decisions made by the AMT and the conventional approach. Consequently, to determine whether the AMT procedure reduced the number of items given to trainees without reducing the quality of the mastery decisions made for those trainees, three criteria were evaluated separately for Test 11 and Test 31 for the AMT and conventional testing procedures at each of the mastery levels:

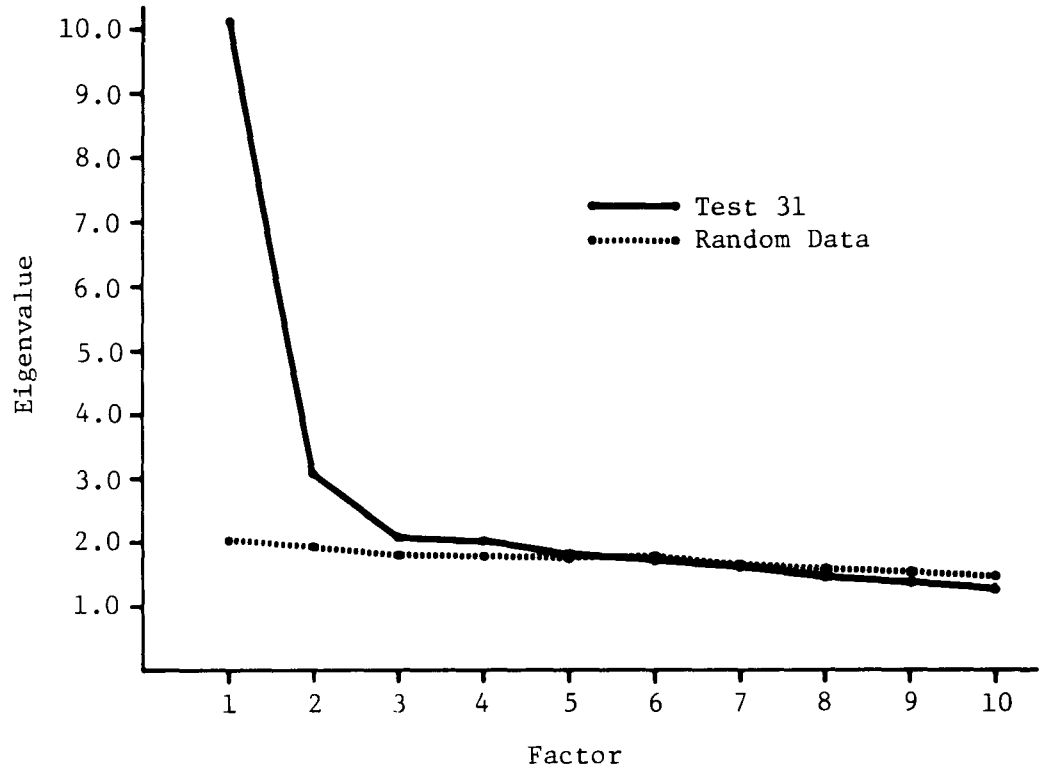
1. The mean number of items administered to trainees,
2. The mean information obtained after all items were administered, and
3. Relationships between mastery decisions made at the termination of the testing by the AMT and conventional procedures.

Figure 3
Eigenvalues of the First 10 Common Factors Extracted From Item Intercorrelations for Test 11 and Test 31 and for Parallel Random-Data Factors

(a) Test 11



(b) Test 31



In addition, to examine the characteristics of the two testing procedures more closely, the mean information obtained from each procedure was plotted for each testing strategy as a function of the achievement level estimate for each mastery level.

Results

Applicability of the ICC Model

Factor analysis. Eigenvalues of the first 10 factors extracted from item intercorrelations for Test 11 and Test 31 and the random data parallel analysis for each test are shown in Appendix Table B-1; these values are plotted in Figure 3. For Test 11 (Figure 3a) the first three factors had higher eigenvalues than their corresponding random-data factors. However, only the first factor differed substantially from the corresponding random-data factor. Thus, for Test 11 it was not unreasonable to infer that only the first factor was a "true" factor underlying trainees' responses, since the eigenvalues of the other factors resembled those of the random factors and the first factor accounted for more than three times the amount of common variance than any other factor.

For Test 31 (Figure 3b) the eigenvalues of the first five factors extracted each exceeded the eigenvalues of their corresponding random factor, but only the first two factors exceeded the random-data values by a substantial amount. The first factor accounted for 20.5% of the common variance extracted by the 10-factor solution, and the second factor accounted for 6.2% of the common variance. No other factor accounted for more than 5% of the variance. These data indicate that there were probably two real factors underlying trainees' responses to Test 31. This two-factor solution might indicate that a multidimensional latent trait model should be postulated to explain trainees' responses to Test 31. However, because the first factor accounted for over three times as much variance as the second factor, the unidimensional model could still be used; data presented by Reckase (1978) indicate that if a dominant first factor exists, items calibrated using a unidimensional model will adequately measure that first factor.

Estimation of the ICC parameters. Tables 1 and 2 show the ICC parameter estimates obtained for each of the items in Test 11 and Test 31, respectively. Of the items in the conventional test, 17% (5 items) from Test 11 were rejected by the parameterization procedure, while 24% (12 items) were rejected for Test 31. These losses are comparable to losses observed during other investigations of achievement tests using this parameterization procedure; Bejar, Weiss, and Kingsbury (1977) lost 22% of their total pool during item parameterization, and Brown and Weiss (1977) lost 13% of their total pool.

For Test 11, values of the a parameter estimates ranged from .63 to 4.69, with a mean of 1.48 and a standard deviation of .98. Values of the b parameter estimates ranged from -2.35 to 1.32, with a mean of -.98 and a standard deviation of 1.01. Values of the c parameter estimates ranged from .00 to .49, with a mean of .27 and a standard deviation of .138.

For Test 31, values of estimates of the a parameter ranged from .63 to 3.42, with a mean of 1.16 and a standard deviation of .65. Values of the b parameter estimates were from -1.86 to 3.18, with a mean of -.58 and a standard deviation of 1.08. The c parameter estimates ranged from .00 to .77, with a

Table 1
ICC Item Parameter Estimates for the Items in Test 11

Item Number	*a* Discrimination	*b* Difficulty	*c* Lower Asymptote
1	--[a]	--	--
2	.81	-.88	.22
3	--	--	--
4	.92	-1.58	.18
5	.66	-1.06	.37
6	.70	-1.18	.36
7	2.75	-1.98	.12
8	1.77	.81	.49
9	1.52	.26	.48
10	--	--	--
11	--	--	--
12	.63	-1.89	.29
13	1.38	-1.64	.31
14	1.70	-1.01	.37
15	1.17	-1.61	.25
16	.67	-1.90	.29
17	1.46	-.74	.27
18	.75	-.93	.12
19	.65	-1.24	.20
20	1.08	-1.71	.36
21	4.69	.98	0
22	2.16	-1.51	.16
23	2.16	-1.55	.19
24	1.32	.56	.30
25	1.21	-1.54	.36
26	--	--	--
27	3.58	-2.35	--
28	1.04	1.32	.46
29	.83	-1.69	.09
30	1.31	-.46	.43

[a]Missing values indicate that the item was rejected by the parameter estimation procedure.

mean of .28 and a standard deviation of .16. For both of these tests the parameter estimates obtained were well within the range established by two earlier studies that examined achievement tests using the same item parameterization method (Bejar, Weiss, & Kingsbury, 1977; Brown & Weiss, 1977).

Examination of the item parameter estimates obtained from Test 11 and Test 31, using Urry's guidelines for a good adaptive testing item pool, indicated the following:

1. For both Test 11 and Test 31, 76% of the items had *a* values exceeding .80, while the average value for both tests exceeded 1.00.
2. The *b* values were fairly widely and evenly distributed between -2.0 and 1.0, but the distribution was rather sparse above 1.0. Considering the small numbers of items in the two item pools, the distribution of the *b* values seems appropriate, though the pools might have been

Table 2
ICC Item Parameter Estimates for the Items in Test 31

Item Number	*a* Discrimination	*b* Difficulty	*c* Lower Asymptote
1	--[a]	--	--
2	.70	-1.40	.33
3	3.39	-1.86	--
4	1.95	3.18	.77
5	.88	-1.78	.37
6	.65	- .82	.14
7	.71	- .68	.39
8	.81	-1.85	.38
9	.66	-1.84	.35
10	--	--	--
11	1.18	-.74	.37
12	--	--	--
13	.95	-.90	.36
14	2.55	-1.39	.01
15	--	--	--
16	.94	-.44	.13
17	1.13	-1.43	.23
18	.92	-.46	.38
19	1.03	-.49	.13
20	.79	.26	.16
21	.80	-1.04	.35
22	1.01	-.65	.15
23	.80	-1.11	.19
24	.79	.98	.27
25	--	--	--
26	1.05	.09	.41
27	.95	-.23	.39
28	1.11	-1.64	.20
29	1.54	-1.56	.14
30	.73	-.44	.11
31	.63	-1.54	.06
32	--	--	--
33	.95	.40	.17
34	1.20	1.13	.45
35	1.07	.45	.27
36	3.42	-1.74	--
37	--	--	--
38	--	--	--
39	--	--	--
40	--	--	--
41	--	--	--
42	--	--	--
43	1.04	-.77	.37
44	1.18	-.49	.39
45	1.03	-.97	.36
46	.74	-1.83	.21
47	1.08	-.56	.38
48	1.02	.80	.37
49	.83	.29	.42
50	1.70	1.06	.33

[a]Missing values indicate that the item was rejected by the parameter estimation procedure.

slightly too easy to meet Urry's second guideline. However, Urry's guidelines were proposed for ability tests for which it is desired to measure precisely across a wide range of ability, whereas the data of this study were from a mastery achievement test for which it was desired to classify students on either side of a mastery level. Thus, the distribution of b values would not be expected to conform with Urry's second recommendation.

3. Fifty-six percent of the items in Test 11 and 47% of the items in Test 31 obtained c estimates below .30. The average c estimate for each test was less than .30.

Thus, in light of Urry's guidelines and the earlier studies, examination of the item parameters obtained indicated that the parameter estimates obtained from Test 11 and Test 31 were similar to those obtained for items which had previously been used to improve achievement measurement; consequently, the items were appropriate for investigating the AMT strategy.

Conversion of the Mastery Level to the ICC Metric

The ICC item parameter estimates for each test were used in Equation 1 to obtain the TCC for each test. Figure 4 shows the resulting TCC for Test 11 (Figure 4a), using item parameters for the 25 items that survived the calibration procedure, and for Test 31 (Figure 4b), based on the 38 items for which parameter estimates were available on that test. Conversion of the proportion correct mastery levels (P=.7, .8, and .9) to the achievement metric (θ) are also shown.

Test 11 had a slightly steeper TCC than did Test 31, reflecting the higher average discrimination of its items. The lower average b level of the Test 11 items (i.e., easier items) is reflected in the fact that the TCC for Test 11 is shifted to the left along the achievement level, or θ, axis in comparison to Test 31. The relatively equal average c parameters for the two tests are reflected in the values of the TCC at θ=-4.0.

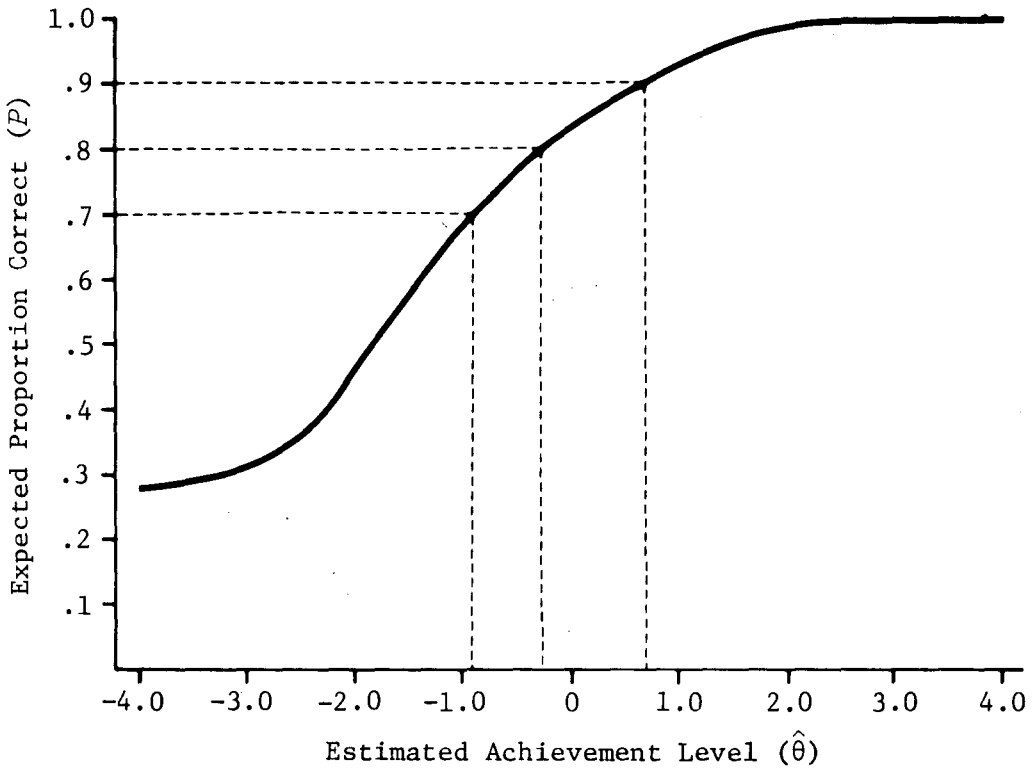
For Test 11 the P=.7 mastery level was converted to θ=-.90 on the achievement metric, the P=.8 mastery level was converted to θ=-.23, and the P=.9 mastery level was converted to θ=.75. For Test 31 the P=.7 mastery level was converted to θ=-.48; the P=.8 level, to θ=.12; and the P=.9 level, to θ=.91 on the achievement metric. It can be seen that for both tests the conversion was nonlinear, reflecting the gain in potential discriminability resulting from consideration of the unique operating characteristics of each item.

Test Length

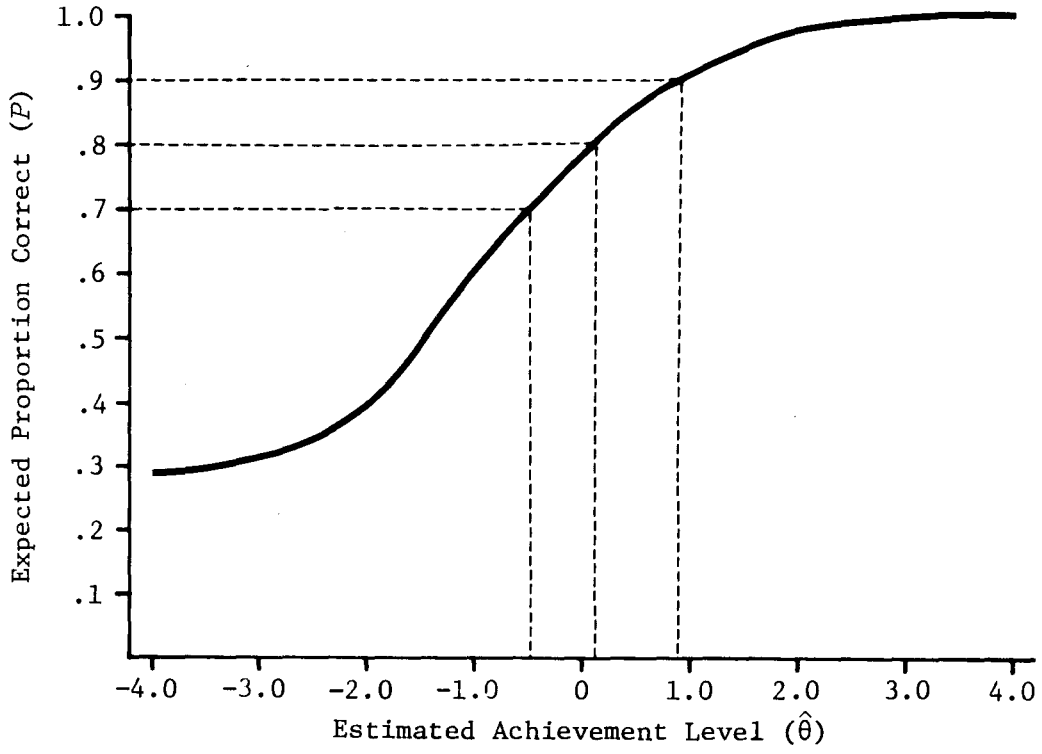
Table 3 shows the mean number of items, the average amount of information obtained from each item administered, and the number of individuals from various subsamples under the AMT and conventional strategies at each of the three different mastery levels. The four subgroups for which these data are presented are (1) the total group of trainees, (2) the groups of trainees declared masters by the relevant testing procedure, (3) the groups of trainees declared nonmasters by the relevant testing procedure, and (4) the groups of trainees for which the AMT procedure made decisions with full confidence (i.e., trainees for whom the mastery level, θ_m, fell outside the 95% confidence interval at some point during the test and terminated the AMT procedure). Frequency distributions of

Figure 4
Test Characteristic Curves for Test 11 and Test 31, with Conversion of Three Mastery Levels (P=.7, .8, and .9) from the Proportion-Correct Metric to the Achievement Metric

(a) Test 11



(b) Test 31



numbers of items administered for each of these subgroups are in Appendix Table B-2 for Test 11 and Appendix Table B-3 for Test 31.

Table 3
Sample Size (N), Mean Test Length ($\bar{L}$), and Mean Information Per Item ($\bar{I}$) for AMT and Conventional (Conv) Test for Tests 11 and 31 at Three Mastery Levels for Total Group and Three Subgroups

Test, Mastery Level, and Testing Strategy	Group											
	Total			Mastery			Nonmastery			High Confidence		
	N	$\bar{L}$	$\bar{I}$	N	$\bar{L}$	$\bar{I}$	N	$\bar{L}$	$\bar{I}$	N	$\bar{L}$	$\bar{I}$
Test 11												
P=.7												
Conv	199	25	.29	172	25	.28	27	25	.39	154	25	.30
AMT	199	12.8	.32	174	12.3	.28	25	16.4	.52	154	9.2	.36
P=.8												
Conv	199	25	.29	135	25	.29	64	25	.30	100	25	.40
AMT	199	17.4	.29	126	17.4	.28	73	17.4	.32	100	9.9	.55
P=.9												
Conv	199	25	.29	43	25	.50	156	25	.23	132	25	.27
AMT	199	13.1	.34	34	20.8	.50	165	11.5	.27	132	7.0	.38
Test 31												
P=.7												
Conv	200	38	.22	127	38	.18	73	38	.28	122	38	.21
AMT	200	21.8	.30	134	19.4	.26	68	26.4	.36	122	11.4	.44
P=.8												
Conv	200	38	.22	73	38	.15	127	38	.26	117	38	.24
AMT	200	23.4	.26	74	27.7	.20	126	20.9	.32	117	13.1	.41
P=.9												
Conv	200	38	.22	27	38	.12	173	38	.23	151	38	.24
AMT	200	14.7	.23	28	38	.12	172	10.9	.30	151	7.2	.40

Total group. For the total group of trainees responding to Test 11, the AMT procedure reduced the average number of items administered ($\bar{L}$) substantially at every mastery level. The minimum reduction in number of items administered that was noted was for the P=.8 mastery level, where test length for the conventional test was 25 items, compared to a mean test length for the AMT procedure of $\bar{L}$=17.4 items; this reduction of 7.6 items represents a minimum test length reduction of 30.4% of the conventional test length. The maximum test length reduction was 48.8% of the conventional test (12.2 items) when a mastery level of P=.7 was used. For the same group of trainees, a gain in the average amount of information ($\bar{I}$) obtained from each item administered was noted for the AMT procedure at the P=.7 and P=.9 mastery levels. The gains in information per item administered were .03 information units (IU), or a 10% increase at the P=.7 mastery level, and .05 IU, or a 17% increase, at the P=.9 mastery level.

For the total group of trainees responding to Test 31, the same two trends were noted. First, test length was reduced with the use of the AMT procedure at each mastery level. The minimum reduction of test length was noted with the

use of the P=.8 mastery level, for which the conventional test length of 38 items was reduced to a mean AMT length of $\bar{L}$=23.4 items--a reduction of 38.4% in mean test length. The greatest reduction in test length was noted for the .9 mastery level at which the mean AMT length was 14.7--a reduction in test length of 61.3%.

The second trend was that the AMT procedure provided more information with each item administered than the conventional test for all mastery levels. The smallest increase in information was .01 IU per item (a 5% increase), for the P=.9 mastery level. The largest gain in the mean information per item was .08 IU (a 36% increase), for the P=.7 mastery level. For mastery levels P=.8 and P=.9, the percent reduction in test length under AMT was greater for Test 31 than that noted for Test 11. The increase in information per item noted for AMT was greater for Test 31 than for Test 11 at all three mastery levels.

Appendix Tables B-2 and B-3 show that test lengths for the AMT procedure for different trainees were quite variable. For most of the trainees, either a very long test (as long as the conventional test) was needed, or a very short test (8 items or less) was sufficient. This U-shaped distribution of test lengths was obtained for both Test 11 and Test 31 across all mastery levels.

Mastery groups. When only those trainees were considered who were judged to be masters for Test 11 at one of the mastery levels by the AMT or the conventional testing procedure, test length reduction was again noted for the AMT procedure at all three mastery levels. For mastery levels P=.7 and P=.8, adaptive tests for those in the mastery group were approximately the same mean length as those for the total group; but for mastery level P=.9 adaptive tests for the mastery group were much longer (20.8 versus 13.1 items on the average). In comparison with the conventional test, for the AMT procedure in the mastery group alone the minimum test length reduction was 4.2 items, or 16.8% of the conventional test length of 25 items, at the P=.9 mastery level; and the maximum test length reduction was 12.7 items, or 50.8% of the conventional test length, at the P=.7 mastery level.

The AMT procedure and the conventional testing procedure provided almost identical mean amounts of information ($\bar{I}$) for items administered to the mastery groups, even though the AMT procedure administered fewer items at each mastery level. However, for these groups interpretation of the differences in mean information ($\bar{I}$) is obscured by the fact that the two different testing procedures gave trainees with different achievement levels mastery status. A clearer comparison of information provided by the two testing procedures is shown below.

For the groups of trainees labeled as masters for Test 31, test-length reduction was observed with the use of AMT for only two of the three mastery levels examined. At the P=.7 mastery level, mean test length was reduced by 18.6 items, or a reduction of 48.9% of the conventional test length, by use of AMT. For the P=.8 mastery level the mean test length was reduced by 10.3 items, or a reduction of 27.1% of the conventional test length. For the P=.9 mastery level the AMT procedure never reached a decision of mastery in less than 38 items, the length of the conventional test.

For Test 31, the AMT procedure resulted in higher mean information per item than the conventional test for the P=.7 mastery level (a difference of

.08 IU per item, or a 44% increase over the conventional test) and the P=.8 mastery level (.05 IU per item higher, a 33% increase). At the P=.9 mastery level the conventional test and the adaptive test administered items with equal average information.

As the mastery level became higher, for both Test 11 and Test 31 there was a trend for greater numbers of items to be administered before a decision of mastery could be made. This resulted from the fact that the higher mastery levels fell above the steepest portion of the TCCs, as is shown in Figure 4. This would imply that the entire conventional test would have more difficulty discriminating among trainees at these mastery levels; consequently, the AMT procedure would have to use more of the items from the conventional test in order to determine whether a trainee was above or below the higher mastery levels. This trend may be clearly seen in Appendix Tables B-2 and B-3. For each test, trainees were placed in the mastery group for mastery level P=.7 with a wide range of test lengths. As the mastery level was raised, trainees were more likely to be declared masters only after a larger number of items were administered, until for Test 31 at the P=.9 mastery level, all those who were declared masters took all of the items in the item pool before the mastery decision was made.

Nonmastery groups. For the trainees who were declared nonmasters for Test 11, using either the adaptive or conventional testing procedures, reductions in test length were observed at every mastery level with the AMT procedure. The smallest reduction in test length, 7.6 items, was observed for the P=.8 mastery level and accounted for 30.4% of the conventional test length. The largest reduction is test length was 13.5 items at the P=.9 mastery level, or 54% of the conventional test length. At each mastery level for Test 11, more mean information was obtained from each item administered to the nonmasters by the AMT procedure than by the conventional procedure. The smallest increase in information per item was .02 IU (a 6.7% increase), for the P=.8 mastery level. The largest increase in mean information was .13 IU (a 33.3% increase) per item, for the P=.7 mastery level.

For the trainees declared nonmasters for Test 31, reductions in mean test length were again noted with the AMT procedure at each mastery level. The minimum mean decrease in test length was 11.6 items, or 30.5% of the conventional test length of 38 items, at the P=.7 mastery level. The maximum reduction in average test length was 27.1 items, or 71.3% of the conventional test length, at the P=.9 mastery level. As the criterion level increased, the number of items needed by the AMT procedure to make the nonmastery decision steadily decreased.

For the nonmastery groups administered Test 31, the mean information per item was higher at each mastery level for the AMT procedure than for the conventional testing procedure. The minimum increase in information was .06 IU (a 23% increase) per item administered, for the P=.8 mastery level; and the maximum increase observed was .8 IU per item (a 28.6% increase), for the P=.7 mastery level.

Across both Tests 11 and 31, there was a tendency for the adaptive test to administer fewer items before making a decision of nonmastery as the mastery level increased. The sole exception to this trend was observed for Test 11 at the P=.8 mastery level, which showed a slight increase in the number of items

administered when compared with the P=.7 mastery level for that test. For both tests a higher mean information was obtained for each item administered by the AMT procedure at each mastery level. No consistent trend was noted in the differences in average information per item across mastery levels for the two tests.

High-confidence groups. The high-confidence groups included only those trainees for whom the AMT procedure terminated with full confidence, i.e., trainees for whom the Bayesian confidence interval failed to include the mastery level at some test length at or before the exhaustion of the items from the conventional test item pool. For Test 11 the AMT procedure terminated with high confidence for a minimum of 50% of the group of trainees, at the P=.8 mastery level. The largest high-confidence group was 77% (N=154) of the total group of trainees, at the P=.7 mastery level.

Test length was reduced considerably by the AMT procedure at all criterion levels for the high-confidence groups. The minimum reduction in mean test length was observed for the P=.8 mastery level and was 15.1 items, or 60.4% of the conventional test length. The largest mean reduction in test length observed was 18 items, or 72% of the conventional test length, at the P=.9 mastery level. Modal test length for the high-confidence groups for Test 11 at all mastery levels was 3 items (see Appendix Table B-2), or only 12% of the length of the conventional test (an 88% reduction). The AMT procedure produced greater mean information per item at each mastery level. The smallest observed increase was .06 IU (a 20% increase) per item administered, for the P=.7 mastery level. The largest mean increase was .15 IU per item (a 37.5% increase), at the P=.8 level. For Test 31 the minimum number of trainees in the high-confidence group was 117, or 58% of the total group, at the P=.8 mastery level. The largest high-confidence group was 151, or 76% of the total trainee group, for the mastery level P=.9.

Test length for the AMT procedure was much shorter than the conventional test at each criterion level. The smallest reduction in mean test length was 24.9 items, or 65.5% of the conventional test length, for the P=.8 mastery level. The largest average reduction in test length was 30.8 items, or 81.1% of the total conventional test length, for the P=.9 mastery level. Similar to Test 11, modal test lengths for Test 31 were quite short: 4 items for the P=.7 mastery level, 5 items for the P=.8 mastery level, and 3 items (for 57% of the high-confidence group) at the P=.9 mastery level.

The AMT procedure produced higher mean information per item than the conventional testing procedure at all mastery levels. The minimum increase in mean information per item was .16 IU (an increase of 66.7% over the mean information provided by the conventional test), for the P=.9 mastery level. The maximum mean information increase that was observed was .23 IU per item (a 112% increase), for the P=.7 mastery level.

For both Test 11 and Test 31 the AMT procedure made confident decisions for between 50% and 77% of the total group at each mastery level. For the trainees in the high-confidence groups, the average adaptive test length ranged from 19% to 39% of the original conventional test length, while modal test lengths were only 8% to 6% of the conventional test length (i.e., over 90% reduction). Also, the adaptive testing procedure resulted in 20% to 119.5% increase in the mean amount of information obtained per item over the conventional test. The increase in mean information per item was greater for Test 31 than for Test 11 at all criterion levels.

Correspondence Between Decisions

Table 4 shows the Pearson product-moment (phi) correlations between the decisions made by the AMT and conventional testing procedures across all three criterion levels for Test 11 and Test 31. The lowest correlation observed was .67, for Test 11 at the P=.9 mastery level. The highest correlation was .97, for Test 31 at the P=.8 mastery level. The correlations between mastery decisions for Test 31 were higher than for Test 11 at all mastery levels. In addition, the average decision variance in common between the two testing procedures was 79% of the total decision variance.

Table 4
Phi Correlations Between Mastery Decisions Made by AMT and Conventional Testing Procedures for Test 11 and Test 31, at Three Mastery Levels

	Mastery Level		
Test	P=.7	P=.8	P=.9
Test 11	.91	.88	.67
Test 31	.93	.97	.94

To examine more completely the correspondence in decisions made by the AMT and conventional procedures, Table 5 shows joint frequency distributions of decisions for the two testing procedures at each of the three mastery levels for Test 11 and Test 31. The lowest level of agreement between the AMT and conventional testing procedures was noted for Test 11 at the P=.9 mastery level, where the two testing procedures agreed for 178, or 89.4% of the 199 trainees tested. The highest level of agreement was 98.5%, for Test 31 at the P=.8 and P=.9 mastery levels. Across both tests and all criterion levels, the two procedures agreed for 95.9% of the trainees tested. For the longer test (Test 31) the two procedures agreed for 97.9% of the trainees, and for the shorter test (Test 11) the two procedures agreed for 94.0% of the trainees.

Table 5
Joint Distributions of Mastery Decisions Made by AMT and Conventional Tests 11 and 31 at Three Mastery Levels

Mastery Level and AMT Decision	Test 11		Test 31	
	Mastery	Nonmastery	Mastery	Nonmastery
P=.7				
AMT Mastery	171	3	126	6
AMT Nonmastery	1	24	1	67
P=.8				
AMT Mastery	125	1	72	2
AMT Nonmastery	10	63	1	125
P=.9				
AMT Mastery	28	6	26	2
AMT Nonmastery	15	150	1	171

Information Functions

Figures 5 and 6 show the information obtained by Conventional Tests 11 and 31, respectively, and adaptive testing procedures as a function of estimated achievement level ($\hat{\theta}$). (Points plotted in these figures are based on mean information obtained from trainees within a plus or minus .1 range around a given $\hat{\theta}$; numerical values of information are shown in Appendix Table B-4.) Figures 5 and 6 each show three adaptive testing information curves--one for each mastery level examined--and one conventional test curve.

Figure 5 shows that Test 11 was poorly designed to make mastery decisions at middle-range mastery levels ($\hat{\theta}$ between -.5 and +.5, or proportion correct of about P=.75 to P=.85), since the test's information was predominantly concentrated at low achievement levels ($\hat{\theta}$<-1.0), with an information spike caused by a single highly discriminating item (Item 28; see Table 1) at about 1.0 on the achievement continuum. Information functions for the AMT strategy at each of the three mastery levels closely approximated the conventional information function in the region near each respective mastery level ($\hat{\theta}$=.8, -.2, -9). In addition, as achievement level moved away from the mastery levels, the AMT information functions fell below the information function for the conventional test, particularly at the lower achievement levels. Further, as the difference between the achievement level and the mastery level increased, the difference in amounts of information used by the AMT procedure and the conventional procedure tended to become larger. However, for the P=.8 mastery level an upturn in the information function occurred below the -1.3 achievement level, and the difference in information between the conventional and adaptive procedure decreased slightly. The same type of upturn was noted for the P=.9 mastery level, for $\hat{\theta}$ levels below -1.1.

Figure 6 shows that for Test 31 the conventional test information function was monotone decreasing within the observed range of trainees' achievement levels. This implies that Test 31 provided its most precise measurement at low achievement levels and that differences between the two testing procedures should be most noticeable at low achievement levels. The AMT information functions for Test 31 in Figure 6 reinforce the trends noted in Test 11 for each of the mastery levels. That is,

1. The AMT information functions each closely approximated the conventional test information function in the region of the achievement continuum near the appropriate mastery level.
2. For achievement levels beyond the region near the mastery level, the AMT information function was lower than the conventional test information function.
3. The difference in information between the AMT and conventional testing procedures was greater for achievement levels further from the specified mastery level, up to a point.
4. At the lower end of the achievement continuum ($\hat{\theta}$<-.5), an increase in the amount of information provided by the AMT procedure was noted for each of the mastery levels examined. The point on the $\hat{\theta}$ continuum at which the upturn was noted was lower for each successively lower criterion level.

For Test 31 one additional result was noted that did not appear in the Test 11 AMT data: For both the P=.8 and P=.9 criterion levels, a final downturn in the

Figure 5
Mean Obtained Information as a Function of Estimated Achievement Level for AMT and Conventional Test 11 at Three Mastery Levels

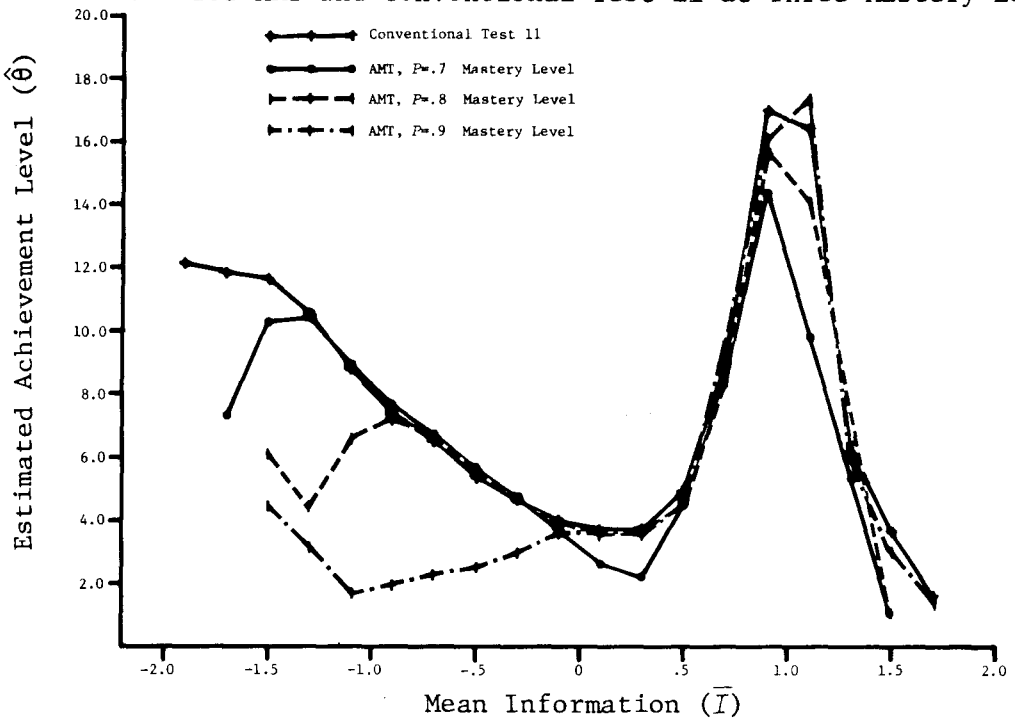
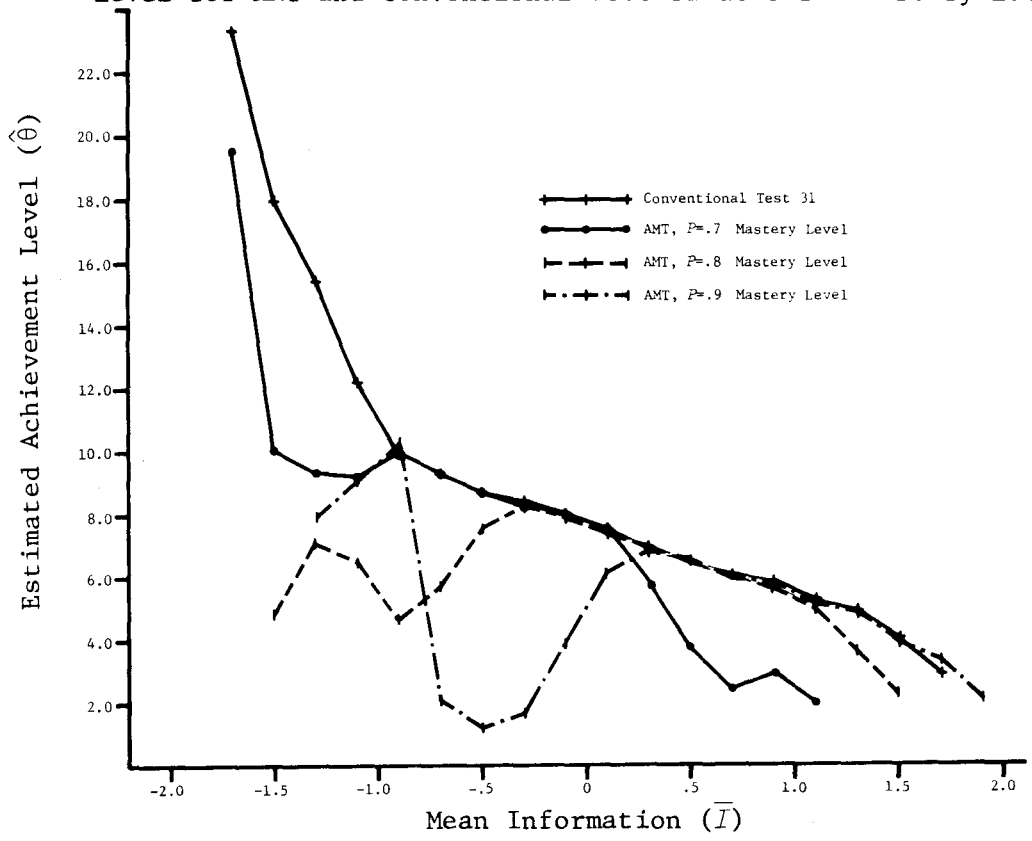


Figure 6
Mean Obtained Information as a Function of Estimated Achievement Level for AMT and Conventional Test 31 at Three Mastery Levels



information functions for the AMT procedure was observed at the lowest obtained $\hat{\theta}$ levels. This implies that the observed upturns in information may have been one side of an information spike, possibly caused by the minimum limit of three items placed on the AMT procedure.

Discussion and Conclusions

The unidimensional three-parameter logistic ICC model was fit to two conventional tests that were previously used to make mastery decisions in a military training course. Data originally gathered during the training course were used to evaluate, in real-data simulation, the efficiency of the proposed adaptive mastery testing (AMT) procedure in terms of the number of items administered, the information obtained, and the degree of agreement between the AMT and conventional testing procedures. The AMT procedure was simulated assuming three different mastery levels, stated in terms of the achievement metric, through the use of the test characteristic curves (TCCs) for the two conventional tests. The results of these simulations indicated that the proposed AMT procedure reduced the number of items administered during the average test, while at the same time making decisions which were very much the same as those made by the conventional testing procedure.

The AMT procedure reduced the average test length for the entire group of trainees by 30% to 61% of the conventional test length. The reductions in test length observed varied across different mastery levels for both of the conventional tests. When specific subgroups of the samples were considered, mean test length reductions of up to 81% of the items in the conventional test were again observed in almost every subgroup examined at each mastery level and for both tests. The only subgroup for which no test length reduction was observed for the AMT strategy was the group passing Test 31 at the highest criterion level (P=.90 correct). For the groups of trainees for which the AMT procedure was able to make high-confidence decisions, AMT mean test lengths were 60% to 81% shorter than the conventional tests across all mastery levels examined. Further, high-confidence decisions were made for 50% to 77% of the trainees at each mastery level.

At each mastery level for each test, agreement was high between the decisions made by the adaptive and conventional testing procedures. The two procedures made the same decision for approximately 96% of the cases across all circumstances. Using the larger item pool (Test 31), the two procedures agreed for about 98% of the cases. The lowest agreement level observed was approximately 89%.

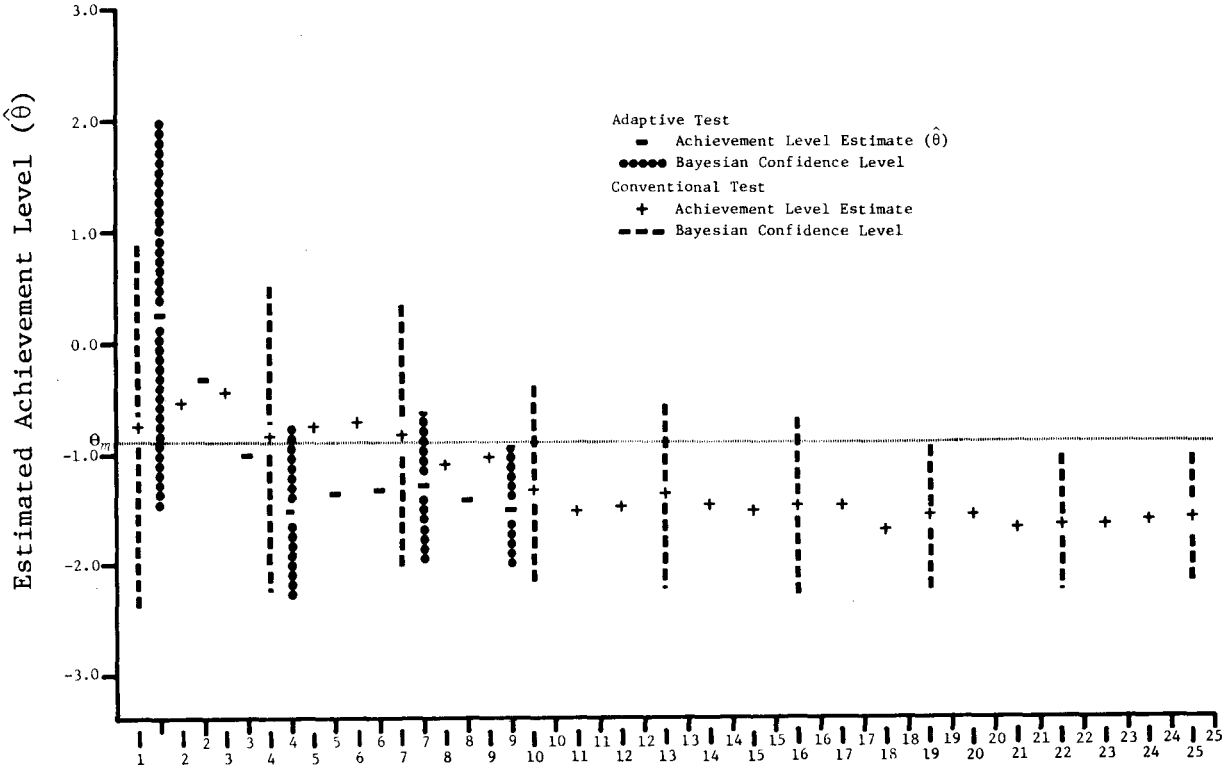
At each mastery level examined, the information functions observed for the adaptive tests closely approximated the information functions obtained for the relevent conventional test at achievement levels close to the mastery level, and fell below the conventional test information functions for more extreme achievement levels. For the achievement levels very different from the mastery level, the difference between the information functions for the two testing procedures reached a maximum; and at the most extreme achievement levels the difference in information decreased slightly.

Thus, the AMT procedure was shown to make mastery decisions very similar to those made by the conventional testing procedure, while administering fewer items, by using the information in the item pool that was available to make high-confidence decisions.

The test-length reduction observed using the AMT procedure may be attributed to two characteristics of the procedure. First, the AMT strategy administered to a trainee only those items which provided the most precise measurement at the trainee's current level of $\hat{\theta}$. Second, the AMT procedure terminated the test as soon as enough information was available to make a decision at a predetermined level of confidence concerning the trainee's mastery level. The termination rule allowed the test to terminate prior to the exhaustion of the item pool, if enough information was available in the items, and the item administration procedure presented the most informative items early in the testing session.

Each of these characteristics of the AMT procedure can be more clearly seen by examination of the Bayesian point estimates and the associated confidence intervals obtained from a trainee's responses after each item administered by the AMT and conventional testing procedures. One such record is shown in Figure 7 for a trainee responding to Test 11. The $\hat{\theta}$ estimates plotted in Figure 7 include 95% Bayesian confidence intervals for the $\hat{\theta}$ estimate after the first item and after every third item administered thereafter for both AMT and conventional procedures (even though the confidence interval was not used for making the mastery decision with the conventional procedure).

Figure 7
Achievement Level Estimates for Trainee 14 after Each Item Administered by AMT and Conventional Testing Procedures for Test 11, with 95% Bayesian Confidence Intervals Indicated after Every Third Item (P=.7 Mastery Level)



It may be seen from Figure 7 that both testing procedures made a nonmastery decision for the trainee (i.e., determined that the trainee's true achievement level fell below the specified mastery level), even though both procedures

estimated the trainee's achievement level as being above the mastery level for the first few items. The conventional test θ estimates were above the mastery level for the first 7 items; the adaptive test θ estimates dropped below the mastery level after only 2 items. The AMT procedure made the mastery decision after administering 9 items, compared with the conventional test length of 25 items. At each test length greater than a single item, the Bayesian confidence interval around the conventional test θ estimate was larger than the confidence interval around the AMT θ estimate. This indicates the greater measurement precision available to the AMT procedure due to the adaptive item administration procedure.

Further, it may be noted in Figure 7 that the conventional test strategy finally resulted in a Bayesian confidence interval that fell completely below the mastery level after 19 items were administered (still over twice the test length of the adaptive test); but since the conventional testing procedure does not terminate even after this high-confidence level is reached, 6 more items were administered before the test ended. This illustrative example showed that the AMT procedure was far more economical than the conventional procedure in terms of test length, due to the adaptive item selection procedure and the use of the Bayesian confidence interval as a termination mechanism.

Additional Advantages of the AMT Strategy

The ICC-based adaptive mastery testing strategy described in this report has several other advantages over conventional testing procedures used to make mastery decisions. As has been demonstrated with these data, use of the ICC metric and related achievement estimation procedures can result in mastery decisions for most trainees (50% to 77%) with known and predetermined levels of confidence. Coupled with appropriate design of mastery testing item pools using ICC concepts, the percentage of high-confidence decisions could be substantially increased until mastery decisions could be made for virtually all students at the same high and predetermined level of confidence. Design of such mastery testing item pools would include a concentration of highly discriminating items around the mastery level, plus sufficient numbers of highly discriminating items elsewhere along the achievement continuum to permit high-confidence decisions to be made for all students. Actual numbers of items required at various discrimination levels could be estimated using Owen's Bayesian scoring procedure and information on the difficulties and discriminations of items to estimate in advance the values of the Bayesian posterior variance (which is used to construct the Bayesian confidence intervals used in the AMT procedure) at the expected levels of θ.

If the mastery testing item pool is not designed in advance to permit high-confidence decisions for each student, the AMT procedure still permits the tester to determine the confidence level of each mastery decision made, even if it is not a high-confidence decision. This can be determined by locating the distance of the mastery level, θ_m, from the student's estimated achievement level, $\hat{\theta}$. This distance can then be treated as a standardized deviation from the mean of a normal distribution, with a variance equal to the estimated posterior variance; and .50 plus the area of the portion of the normal distribution included in that deviation will then give the confidence level for a given mastery decision for that student. In this way, a confidence level for the mastery decision can be attached to each such decision. As a result, instructional decisions based on lower confidence level mastery decisions can be made more tentatively.

A further advantage of the ICC-based AMT strategy is that it can be extended to the multiple-content area mastery testing problem with further savings in test administration time. In many training environments, it is desirable to measure mastery on a number of learning objectives at the same point in time. Using conventional testing procedures to measure mastery on 6 objectives, for example, the student would have to take 6 different tests with a fixed number of items, for a potential total of over 100 items. However, since the AMT strategy utilizes the same item selection and scoring procedures that Brown and Weiss (1977) used in their intercontent branching adaptive testing strategy, the AMT strategy can operate in the same fashion; all that differs is the intrasubtest termination rule. Thus, in the multicontent branching AMT strategy, the achievement level estimates used to make the mastery decisions in each of a number of content-based mastery tests would be used to serve as entry points for beginning testing (using appropriate multiple regression equations) in subsequent mastery tests in the battery. If there is any correlation between mastery decisions made on the separate subtests, the use of an intercontent branching AMT should result in substantial additional savings in testing time over that obtained by use of the AMT strategy in each subtest separately.

The AMT procedure described above, or an improved version, should thus be extremely useful in a training sequence in which many subject areas are taught and tested within a short time, thus putting a premium on testing time. A self-paced instructional setting in which a student is given more than one attempt to demonstrate mastery of a content area with a single test may also benefit from an AMT procedure that would allow students to take different items on each attempt, thus avoiding the problem of students merely "learning" the test, without learning the subject matter.

The AMT procedure should be tested in an actual classroom situation. Further research should also be conducted to determine whether conventional mastery testing or the AMT procedure result in mastery decisions which more accurately predict external performance criteria.

References

Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, January 1979. (NTIS No. AD A067752)

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A047495)

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828)

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

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A046062)

Davis, F. B., & Diamond, J. J. The preparation of criterion-referenced tests. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.), Problems in criterion-referenced measurement. Los Angeles, CA: UCLA Graduate School of Education, Center for the Study of Evaluation, 1974.

Ferguson, R. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction (Doctoral Dissertation, University of Pittsburgh, 1970). Dissertion Abstracts International, 1970, 30, 3856A. (University Microfilms No. 70-4530)

Fisher, R. A. Contributions to mathematical statistics. New York, NY: John Wiley & Sons, 1950.

Glaser, R., & Klaus, D. J. Proficiency measurement: Assessing human performance. In R. M. Gagné (Ed.), Psychological principles in system development. Chicago, IL: Holt, Rinehart, & Winston, 1962.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 1978, 48, 1-48.

Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 1965, 30, 179-185.

Kingsbury, G. G., & Weiss, D. J. Relationships among achievement level estimates from three item characteristic curve scoring methods (Research Report 79-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1979.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Lord, F. M. Discussion. In W. A. Gorman (Chair), Computers and testing: Steps toward the inevitable conquest (PS-76-1). Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center, September 1976. (NTIS No. PB-261-694)

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

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (NTIS No. AD A022964)

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences. New York, NY: McGraw-Hill, 1970.

Owen, R. J. A Bayesian approach to tailored testing (Research Bulletin 69-92). Princeton, NJ: Educational Testing Service, 1969.

Popham, W. J. (Ed.), Criterion-referenced measurement--an introduction. Englewood Cliffs, NJ: Educational Technology Publications, 1971.

Popham, W. J., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement, 1969, 34 (4, Pt. 2, Monograph No. 17).

Urry, V. W. A five year quest: Is computerized adaptive testing feasible? In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Research and Development Center, PS-75-6). Washington, DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-000-00940-9)

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

Vale, C. D., & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1975. (NTIS No. AD A018758)

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376)

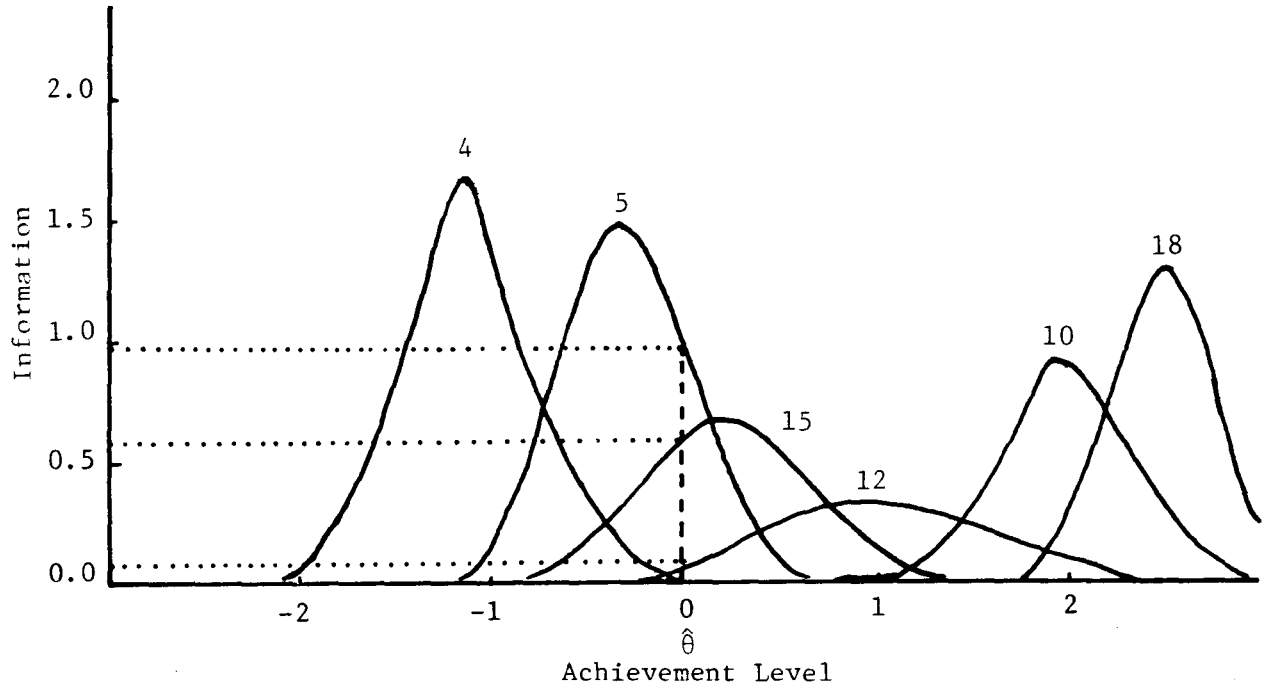
Appendix A

Illustration of MISS Procedure for Choosing Items for AMT

The essential characteristics of the adaptive testing strategy employed in this study have been described in previous sections. However, to understand the method more completely, it is helpful to see the results of its application with an actual testee.

Figure A-1 shows estimated item information curves for six items from Test 1. (There would probably be many items in the test, but only six were chosen to simplify the illustration.) The height of the information curve at a given achievement level ($\hat{\theta}$) indicates the amount of information provided by the item. Most of the items are fairly "peaked"; that is, they provide information over a relatively narrow range of the achievement continuum. While the information curves overlap to some degree, different items provide different amounts of information at a given point on the achievement continuum. The guiding principle for the adaptive procedure is to administer the item which provides the most information at the current achievement estimate ($\hat{\theta}$).

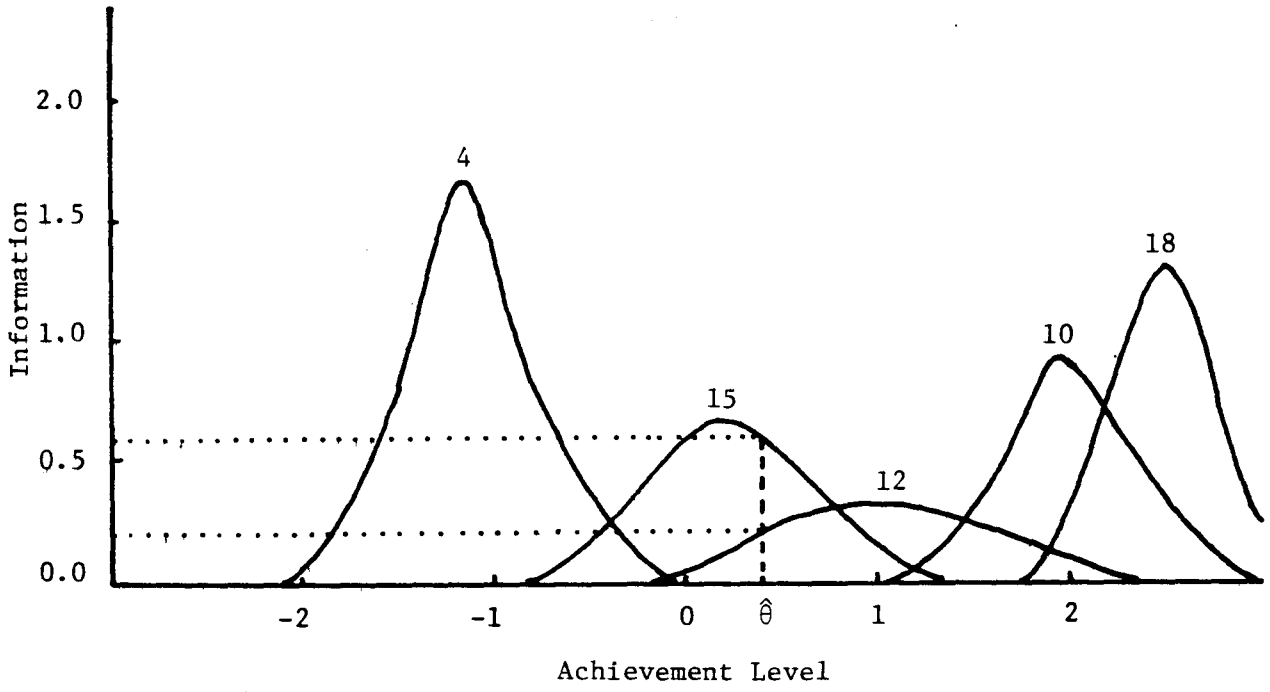
Figure A-1
Estimated Item Information Curves for Six Items from Test 1



For a testee beginning Test 1, the initial achievement estimate was $\hat{\theta}=0$; this is shown by the vertical dashed line in Figure A-1. Of the six items in the example, only three items had essentially nonzero information values at $\hat{\theta}=0$; these values, shown by the horizontal dotted lines in Figure A-1, were .95 for Item 5, .60 for Item 15, and .10 for Item 12. Applying the rule that the item selected is the one which provides the most information at the current $\hat{\theta}$, Item 5 would be selected for administration.

Figure A-2 shows the revised value of $\hat{\theta}$=.46 derived from the Bayesian scoring routine, assuming that a correct answer was given to Item 5. The confidence interval surrounding this $\hat{\theta}$ is assumed to contain the mastery level, so testing would continue. The information curve for Item 5, which was already administered, is not shown in Figure A-2. At the new value of $\hat{\theta}$, only Items 15 and 12 provide significant values of information. Since Item 15 has an information value of .60 and Item 12 has a value of .20, Item 15 would be selected as the second item to be administered to this testee.

Figure A-2
Estimated Item Information Curves for Five Items from Test 1



Assuming that the testee had correctly answered Item 15, the value of $\hat{\theta}$ would increase to .92. The confidence interval around this new $\hat{\theta}$ still contains the arbitrary mastery level, so testing would continue. At $\hat{\theta}$=.92, only Item 12 would provide significant amounts of information, and it would be administered next. Thus, at each step during the testing procedure, the item which provides the most information concerning the testee's current level of $\hat{\theta}$ is administered. In a larger item pool, testing would continue in this fashion until it was possible to make a mastery decision with a prespecified level of confidence, at which point the test would terminate.

Appendix B

Supplementary Tables

Table B-1
Eigenvalues of the First 10 Common Factors Extracted from Item Intercorrelations for Test 11 and Test 31, and for Parallel Random-Data Factors

	Test 11		Test 31	
Factor	Real Data	Random Data	Real Data	Random Data
1	6.14	1.75	10.23	2.04
2	1.85	1.61	3.08	1.90
3	1.60	1.52	2.10	1.84
4	1.41	1.51	2.06	1.82
5	1.38	1.50	1.82	1.76
6	1.30	1.39	1.68	1.72
7	1.24	1.33	1.58	1.62
8	1.16	1.28	1.49	1.58
9	1.15	1.25	1.38	1.56
10	.97	1.20	1.31	1.48

Table B-2
Frequency Distributions of Number of Items Administered by AMT Procedure from Test 11 by Mastery Subgroup for Each Mastery Level (P=.7, .8, and .9)

Number of Items Administered	Group											
	Total			Mastery			Nonmastery			High Confidence		
	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9
3	43	54	45	39	39		4	15	45	43	54	45
4	1	1	24				1	1	24	1	1	24
5	36	1	10	36				1	10	36	1	10
6	1		3				1		3	1		3
7	10	3	17	10		8		3	9	10	3	17
8	13	2	2	13	1			1	2	13	2	2
9	3		2				3		2	3		2
10			1						1			1
11	7		6	7					6	7		6
12	1		3				1		3	1		3
13	3	2		3				2		3	2	
14	1	2	1	1					1	1	2	1
15		3	2		2			1	2		3	2
16	1	1	1	1				1	1	1	1	1
17	1	7	4		6		1	1	4	1	7	4
18	7		2	7					2	7		2
19	4	6	1	1			3	6	1	4	6	1
20	4			4						4		
21		1	3					1	3		1	3
22		2	2		2				2		2	2
23	1		3	1					3	1		3
24	7	9		7	8			1		7	9	
25	55	105	67	44	68	26	11	37	41	10	6	

Table B-3
Frequency Distributions of Number of Items Administered by AMT Procedure From Test 31 by Mastery Subgroup for Each Mastery Level (P=.7, .8, and .9)

Number of Items Administered	Group											
	Total			Mastery			Nonmastery			High Confidence		
	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9	P=.7	P=.8	P=.9
3	7	7	86				7	7	86	7	7	86
4	27	1	11	27				1	11	27	1	11
5	4	15	7				4	15	7	4	15	7
6	8	6	6	7			1	6	6	8	6	6
7	10	10	4	10	7			3	4	10	10	4
8	6	4	2	5			1	4	2	6	4	2
9	6	3	1	5			1	3	1	6	3	1
10	7	6	7	5	4		2	2	7	7	6	7
11	1	10	4		3		1	7	4	1	10	4
12	1	4	3	1	1			3	3	1	4	3
13	2	8	2	2				8	2	2	8	2
14	5	1	2	4			1	1	2	5	1	2
15	3	2		3				2		3	2	
16	5	2	2	5				2	2	5	2	2
17	2	6		1	5		1	1		2	6	
18	3	6		1	5		2	1		3	6	
19	4	4	1	1	3		3	1	1	4	4	1
20	2	1		1			1	1		2	1	
21	3	4		2			1	4		3	4	
22			2						2			2
23	5	2		4			1	2		5	2	
24	1	5	1	1	2			3	1	1	5	1
25	1	1		1				1		1	1	
26												
27	1	1		1	1					1	1	
28		1	1		1				1		1	1
29	3	1			1		3			3	1	
30	1		2				1		2	1		2
31	2	1		1			1	1		2	1	
32		1						1			1	
33		1	1		1				1		1	1
34			2						2			2
35	1			1						1		
36	1	2	2				1	2	2	1	2	2
37		1	2					1	2		1	2
38	78	83	49	43	40	28	35	43	21			

Table B-4

Mean Information ($\bar{I}$) Obtained by AMT and Conventional Testing Procedures for Tests 11 and 31 At Three Mastery Levels (P=.7, .8, and .9) for Trainees with Various Achievement Level Estimates ($\hat{\theta}$), and Number of Trainees (N) at Each Achievement Level

$\hat{\theta}$ Range		Test 11								Test 31							
		Conventional		AMT (P=.7)		AMT (P=.8)		AMT (P=.9)		Conventional		AMT (P=.7)		AMT (P=.8)		AMT (P=.9)	
Lo	Hi	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N
-2.000	-1.800	12.15	1														
-1.799	-1.600	11.88	5	7.42	2					23.44	2	19.58	1				
-1.599	-1.400	11.59	3	10.27	7	6.20	7	4.47	4	18.08	1	10.10	8	4.82	4		
-1.399	-1.200	10.58	3	10.46	4	4.46	12	3.19	7	15.53	10	9.36	12	7.23	10	8.00	6
-1.199	-1.000	8.87	10	8.93	6	6.56	4	1.60	4	12.31	9	9.19	11	6.57	8		
-.999	-.800	7.47	4	7.68	6	7.17	11			10.11	18	10.07	7	4.71	22	10.36	4
-.799	-.600	6.61	12	6.63	10	6.69	10	2.34	47	9.30	21	9.35	21	5.80	28	2.17	25
-.599	-.400	5.41	10	5.51	9	5.56	10	2.53	25	8.79	15	8.78	15	7.61	12	1.24	24
-.399	-.200	4.65	14	4.73	9	4.68	18	3.05	20	8.46	19	8.43	17	8.40	17	1.70	58
-.199	.000	4.04	21	3.79	16	4.03	18	3.68	10	8.06	18	8.06	15	8.04	17	3.96	22
.001	.200	3.73	15	2.69	24	3.72	20	3.64	15	7.58	12	7.57	4	7.51	12	6.10	11
.201	.400	3.72	19	2.21	47	3.71	9	3.73	13	6.96	15	5.83	15	6.95	12	6.89	4
.401	.600	4.91	19	4.69	4	4.60	9	4.56	12	6.48	16	3.88	23	6.50	14	6.51	8
.601	.800	8.86	22	8.80	11	8.37	16	9.41	9	6.05	7	2.60	18	6.00	6	6.00	6
.801	1.000	17.01	16	14.40	1	15.81	9	16.09	9	5.78	6	2.93	5	5.61	9	5.74	6
1.001	1.200	16.45	9			13.94	3	17.33	7	5.24	11	1.97	27	4.67	13	5.18	10
1.201	1.400	6.35	3					6.10	2	4.89	6			3.51	8	4.86	6
1.401	1.600	3.70	3	1.29	39	1.29	39	3.16	3	4.05	4			2.20	7	3.84	3
1.601	1.800	1.63	5					1.41	8	2.94	5					3.41	1
1.801	2.000															2.13	5

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

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

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 CDR Robert S. Kennedy
Naval Aerospace Medical and Research Lab
Box 29407
New Orleans, LA 70189

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION & TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs (Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc (OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

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

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 LCOL Gary Bloedorn
Training Effectiveness Analysis Division
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, NM 88002

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. 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 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

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

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

Marines

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

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM
Ft. 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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

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

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Warner Birice
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Educational Psychology
University of Illinois
Champaign, IL 61820

1 Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairmar
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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

1 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

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

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

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

1 Dr. Robert Tsutakawa
Dept. of Statistics
University of MIssouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.

79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.(NTIS No. AD A068176)

79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979. (NTIS No. AD A067928)

79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979. (NTIS No. AD A067752)

78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978. (NTIS No. AD A064352)

78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.

78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978. (NTIS No. AD A062797)

78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.

78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978. (NTIS No. AD A059436)

77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977. (NTIS No. AD A047495)

77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977. (NTIS No. AD A046062)

77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977. (NTIS No. AD A044828)

77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977. (NTIS No. AD A041090)

77-3. Accuracy of Perceived Test-Item Difficulties. May 1977. (NTIS No. AD A041084)

77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.

77-1. Applications of Computerized Adaptive Testing. March 1977. (NTIS No. AD A038114)
Final Report: Computerized Ability Testing, 1972-1975. April 1976. (NTIS No. AD A024516)

76-5. Effects of Item Characteristics on Test Fairness. December 1976. (NTIS No. AD A035393)

76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976. (NTIS No. AD A027170)

76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976. (NTIS No. AD A028147)

76-2. Effects of Time Limits on Test-Taking Behavior. April 1976. (NTIS No. AD A024422)

76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (NTIS No. AD A022964)

75-6. A Simulation Study of Stradaptive Ability Testing. December 1975. (NTIS No. AD A020961)

75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (NTIS No. AD A018675)

75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975. (NTIS No. AD A018758)

75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (NTIS No. AD A013185)

75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975. (NTIS No. AD A007572)

75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (NTIS No. AD A006733).

74-5. Strategies of Adaptive Ability Measurement. December 1974. (NTIS No. AD A004270)

74-4. Simulation Studies of Two-Stage Ability Testing. October 1974. (NTIS No. AD A001230)

74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (NTIS No. AD 783553)

74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (NTIS No. AD 781894)

74-1. A Computer Software System for Adaptive Ability Measurement. January 1974. (NTIS No. AD 773961)

73-3. The Stratified Adaptive Computerized Ability Test. September 1973. (NTIS No. AD 768376)

73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.

73-1. Ability Measurement: Conventional or Adaptive? February 1973. (NTIS No. AD 757788)

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.

Copies of these reports are available, while supplies last, from:

Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

EFFICIENCY OF AN ADAPTIVE INTER-SUBTEST BRANCHING STRATEGY IN THE MEASUREMENT OF CLASSROOM ACHIEVEMENT

Kathleen A. Gialluca
and
David J. Weiss

RESEARCH REPORT 79-6
NOVEMBER 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Air Force Office of Scientific Research, Army Research Institute, Defense Advanced Research Projects Agency, and Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Kathleen A. Gialluca and David J. Weiss		8. CONTRACT OR GRANT NUMBER*(s)* N00014-79-C-0172
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.: RR042-04 T.A.: RR042-04-01 W.U.: NR150-433
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE November 1979 13. NUMBER OF PAGES 44
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 in 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

This research was supported by funds from the Air Force Office of Scientific Research, Army Research Institute, Defense Advanced Research Projects Agency, and Office of Naval Research, and monitored by the Office of Naval Research.

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

testing	sequential testing	programmed testing
achievement testing	branched testing	response-contingent testing
computerized testing	individualized testing	automated testing
adaptive testing	tailored testing	item characteristic curve theory
		latent trait test theory

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

A real-data simulation was conducted to investigate the efficacy of an adaptive testing strategy designed for achievement test batteries applied to a classroom achievement test. This testing strategy combined adaptive item selection routines both within and between the subtests of the test battery. Comparisons were made between the conventionally administered tests and the simulated adaptive tests in terms of test length, psychometric information, and correlations of achievement estimates. Design of the study also permitted (1) separation of the effects of the adaptive intra-subtest item selection

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

procedure and inter-subtest branching, (2) evaluation of the effects of different intra-subtest termination criteria, (3) use of classical regression equations and regression equations corrected for errors of measurement in the predictors, and (4) cross-validation stability of the inter-subtest branching regression predictions. Data consisted of the responses from 1,600 students to classroom-administered final exams in a general biology course at the University of Minnesota.

Total test length was reduced from 16% to 30% using the adaptive intra-subtest item selection strategy with a variable termination criterion that omits those items providing little information to the measurement process. Subtest-length reductions ranged from about 8% to 62%. Total test length was reduced another 1% to 5% (with subtest-length reductions of up to 53%) upon the addition of an inter-subtest branching strategy that utilized regression equations with prior information concerning a student's performance.

Reductions in subtest length were accomplished with virtually no loss in psychometric information. Correlations between the Bayesian achievement estimates from the adaptive and conventional tests were uniformly high, typically r=.90 and higher. Results showed that the use of the corrected regression equations did little to improve the performance of the inter-subtest branching; although the multiple correlations for the corrected equations were higher, both the information curves and correlations of achievement estimates were generally lower. Cross-validation results indicated that the procedure can be used in different samples from the same population.

Results from this study generally supported the generality of this adaptive testing strategy for reducing achievement test length with no adverse impact on the quality of the measurements. Suggestions are made for further research with this testing strategy.

CONTENTS

CONTENTS, continued

Acknowledgments

Data utilized in this report were obtained from volunteer students in General Biology, Biology 1-011, at the University of Minnesota, during fall quarter 1977 and winter quarter 1978; appreciation is extended to these students for their participation in this research. The cooperation of Kathy Swart and Norman Kerr of the General Biology staff in providing access to the students, as well as their encouragement and contributions of time and ideas to this research program, are deeply appreciated.

Technical Editor: Barbara Leslie Camm

Efficiency of an Adaptive Inter-Subtest Branching Strategy In the Measurement of Classroom Achievement

The development of adaptive testing technology has traditionally taken place within the context of ability measurement. Indeed, much of the adaptive testing research has been concerned with the application of the various adaptive testing strategies to the measurement of a single unidimensional ability domain (e.g., Betz & Weiss, 1974, 1975; Larkin & Weiss, 1974, 1975; Lord, 1977; McBride & Weiss, 1976; Urry, 1977; Vale & Weiss, 1975; Weiss, 1973). More recently, Bejar and Weiss (1978); Bejar, Weiss, and Gialluca (1977); Bejar, Weiss, and Kingsbury (1977); and Kingsbury and Weiss (1979) have demonstrated the applicability of these unidimensional adaptive testing strategies to the measurement of classroom achievement. Frequently, however, achievement tests include items drawn from several distinct content areas. Hence, the assumption of unidimensionality of the entire set of items constituting an achievement test may be untenable, and the application of unidimensional testing strategies inappropriate.

Although Reckase (1978) has shown that the first factor of a multidimensional achievement test will be related to the item characteristic curve (ICC) item parameter estimates from the three-parameter ICC model, in many cases the first factor will account for only a small portion of the common variance of the achievement test items, and even smaller portions of the total variance of the test. Thus, application of a unidimensional ICC model to a multidimensional achievement test will result in achievement level estimates that reflect achievement on only a small subset of course content. In addition, the diagnostic information regarding a student's performance on specific course content areas is lost to both student and instructor by measuring achievement on only one dimension.

In an attempt to design an adaptive testing strategy that would reduce testing time, yet retain the capability of providing students and instructors with scores on the separate subtests in an achievement domain, Brown and Weiss (1977) proposed a testing strategy specifically designed for achievement test batteries that are composed of multiple content areas. It included provisions for adaptive branching between subtests as well as for adaptive item selection within subtests, in an attempt to adapt the test battery to each examinee most efficiently. Brown and Weiss (1977) applied the combined inter-subtest and intra-subtest adaptive strategy in a real-data simulation using a military achievement test battery. They observed a mean reduction in test battery length of nearly 50%, accompanied by a minimal loss in psychometric information.

Purpose

The present study investigated the efficacy of this adaptive testing strategy when it was applied to a classroom achievement test in a different kind of testing environment. Further, this study evaluated the relative contributions of the intra-subtest item selection and inter-subtest branching strategies in

terms of

1. The number of items administered in each subtest of the battery and in the test as a whole,
2. Reduction in test length when compared to the length of a conventionally administered examination,
3. Correlations between achievement estimates derived from the adaptive strategies with those obtained from the conventional examination, and
4. Effects of adaptive administration on psychometric information.

In addition, this study included an investigation of the effects of using the adaptive inter-subtest branching strategy developed from one set of data on a different data set, using a double-cross-validation design.

METHOD

Procedure

Test Items and Subjects

Real-data simulation techniques were applied to the item responses of 800 students who were administered the final examination in General Biology, Biology 1-011, an introductory lecture and laboratory class at the University of Minnesota, during the fall academic quarter of 1977, and to the responses of another 800 biology students from winter quarter of 1978.

Each of these final examinations was 110 items long and was administered conventionally by paper and pencil at the end of the academic quarter. However, each student was directed to answer only 100 of the questions and was free to omit any 10 items of his/her choice. Additionally, only the responses to those items from five content areas--Chemistry, Cell, Energy, Reproduction, and Ecology--were used for this study. The numbers of items in each content area differed slightly across the two quarters; the distribution of items across content areas for the two quarters is shown in Table 1. Each of these five content areas formed a subtest used for the branching strategy discussed below.

Item Parameterization

Items were parameterized within content areas using Urry's (1976) ESTEM computer program for latent trait item parameterization employing the three-parameter logistic model. This program provides estimates of the ICC item discrimination (a), item difficulty (b), and lower asymptote (c) parameters.

Urry's item parameterization program calculates item parameter estimates using a two-stage procedure. In the first stage, initial item parameter estimates are determined for all items. However, these initial item parameter estimates are not reported for an item if one or more of the following conditions holds: (1) $a < .80$, (2) $b < -4.00$, (3) $b > 4.00$, or (4) $c > .30$. In the second stage, item parameters are recomputed for all items that are not excluded by the criteria applied in the first stage. In this stage, item parameter estimates are reported without restrictions (e.g., c may be greater than .30 for some items in the second stage) for all items not excluded in the first stage.

The items were parameterized at the peak of training; that is, items in each content area were parameterized using test data obtained soon after in-

struction in that content area took place. Items in content areas Chemistry, Cell, and Energy were parameterized at the time of Midquarter 1 (MQ1), and items in content areas Reproduction and Ecology were parameterized at the time of Midquarter 2 (MQ2). Item parameter estimates were obtained from classroom examination data from winter quarter of 1976 through spring quarter 1977. The minimum sample size for parameter estimation for any one item was 844; most item parameter estimates were based on data from 1,000 to 2,000 students.

Conventional Test

A conventionally administered test was used for comparison with the adaptive testing strategies. The subtests were administered in the same order for both the conventional and adaptive strategies. In the conventional test all items within each subtest were administered sequentially, with all students taking all the items, and all items were administered in the same order. There was, then, no differential entry point for the subtests when administered conventionally. Bayesian scoring (Owen, 1975) was used for each of the conventional subtests, using a mean of 0.0 and a prior variance of 1.0 as the initial prior estimate of the Bayesian score for each subtest.

Adaptive Tests

As in the Brown and Weiss (1977) study, an adaptive testing strategy utilizing both inter-subtest adaptive item selection and intra-subtest branching was used, in conjunction with a variable termination criterion. This was done in order to reduce to a minimum the number of items administered to each student, while causing minimal change in the measurement characteristics of the whole test.

As in the conventional test, a Bayesian achievement estimate ($\hat{\theta}$) was obtained for each student after the administration of every item. Item selection within each subtest was based on the concept of item information as described by Birnbaum (1968). Items were selected within a subtest for each student by computing the value of item information for every unadministered item at the current level of $\hat{\theta}$ for that student. The item selected for administration was the item that had the highest item information value at that level of $\hat{\theta}$; once an item was administered to a student, it was eliminated from the subtest pool of available items for that student. The selected item was administered, the student's response was scored, and a new θ estimate was obtained. Then a new item was selected, and the procedure was repeated.

Testing continued within each subtest until one of the following conditions occurred: (1) all the items within the subtest pool were administered; or (2) no item remaining in the pool provided information at the current level of $\hat{\theta}$ that exceeded some predetermined small amount of information. Two such values of information were used in this study: .01 and .05. Further detail regarding item selection and achievement estimation can be found in Brown and Weiss (1977).

Inter-Subtest Branching

Subtest ordering. Following the proposal by Brown and Weiss (1977), linear multiple regression was used to determine the order of administration of the subtests. Brown and Weiss, however, ordered subtests based on the linear regres-

sion of number-correct scores. In this study a Bayesian achievement estimate, using an assumed normal prior distribution with a mean of 0.0 and a variance of 1.0, was calculated for each student on each of the five subtests of the final examination. These five scores were then intercorrelated, and their intercorrelation matrix was used as the basis for inter-subtest branching. This procedure was used for the data from each of the two academic quarters separately.

The highest bivariate correlation was selected from this intercorrelation matrix (for each quarter), and one of the two subtests was arbitrarily designated to be administered first; the other was administered second. Multiple correlations were then computed using these two subtests as predictor variables and each of the other subtests, in turn, as the criterion variable. The subtest having the highest multiple correlation with the first two subtests was designated as the third test to be administered. This procedure was repeated to select the fourth subtest to be administered, selecting that subtest which had the highest multiple correlation with the previous three subtests. This process was continued until all five subtests were ordered and was repeated separately for each of the two quarters.

Differential subtest entry points. After administration of the first subtest, each student's entry points for the second and subsequent subtests were differentially determined. For the first subtest each student's prior achievement level was assumed to be $\hat{\theta} = 0.0$. That is, it was assumed that the student's achievement level was at the mean of the estimated θ distribution, since there was no previous information to indicate otherwise. The initial item administered from the first subtest was that item providing the most information at $\hat{\theta} = 0.0$; hence, all students began the first subtest with the same test item.

The entry point into the item pool for the second subtest was determined from the bivariate regression of scores from Subtest 2 on Subtest 1 and the student's $\hat{\theta}$ at the end of Subtest 1 ($\hat{\theta}_1$). The value of $\hat{\theta}_1$ for each student was entered into the bivariate regression equation for predicting the second subtest score from the score on the first subtest. This yielded an estimate for that student's score on Subtest 2, which was then used as the initial Bayesian prior $\hat{\theta}$ for intra-subtest item selection in Subtest 2. The item that provided the most information at this predicted level of θ was administered as the first item in the second subtest. The squared standard error of estimate from the bivariate regression equation was used as an estimate of the initial Bayesian prior variance of this entry-level achievement estimate.

Determination of the entry point for the third and subsequent subtests was simply a generalization of the method used for the second subtest. In general, the student's final achievement level estimates from all n previously administered subtests were entered into the multiple regression equation for predicting the next (n + lst) subtest score from scores on the previous n subtests. This predicted achievement level estimate was used as the initial Bayesian prior $\hat{\theta}$ for intra-subtest branching within that subtest. The squared standard error of estimate from each regression was used as the initial Bayesian prior variance for each subtest.

Corrected regression equations. In addition to the classical multiple regression equations, a second set of equations was used to determine entry-level achievement estimates for each subtest. This second set of equations was applied to the data from fall and winter final exams in exactly the same manner as described above; the only difference between the two procedures was in the

way the equations were obtained. The results from use of the two kinds of regression equations were then compared.

The use of the second set of regression equations was studied because classical regression techniques were somewhat inappropriate for this set of data. In the general linear model of regression, the expected value of the dependent variable y is expressed as the "best" (in the least squares sense) weighted sum of p independent variables $x_i (i=1, \ldots, p)$. It is assumed that y is randomly distributed with n independent observations $y_j (j=1, \ldots, n)$, with common variance σ^2, and that the independent variables x_i are measured without error (Neter & Wasserman, 1974).

However, the original Bayesian $\hat{\theta}$ values used in this regression, obtained for each subtest of the final exam, were not measured without error. Indeed, for each of these Bayesian estimates, there was a corresponding value for the Bayesian posterior variance, which can be interpreted as an index of the variation inherent in the estimate itself. Hence, any classical regression procedure using these estimates is somewhat in error.

Lawley and Maxwell (1973) and Maxwell (1975) have discussed the effects such errors have on the regression equation and the multiple correlation coefficient. In their discussions, the general linear equation is expressed as

$$y_j = \alpha + \beta_1 (x_{j1} - \bar{x}_1) + \ldots + \beta_p (x_{jp} - \bar{x}_p) + e_j, \qquad [1]$$

where

α is a constant;
β's are the partial regression coefficients;
$\bar{x}_i$ is the mean of x_{ji} over all j; and
e_j is the random error of measurement in y_j.

The estimation equation, found by the method of least squares (where $\Sigma_j e_j^2$ is minimized), can be written as

$$\hat{y}_j = \bar{y}_j + \hat{\beta}_1 (x_{j1} - \bar{x}_1) + \ldots + \hat{\beta}_p (x_{jp} - \bar{x}_p), \qquad [2]$$

where $\bar{y}_j$ is the mean of the n observations of $y_j (j = 1, \ldots, n)$ and $\hat{y}_j$ is the predicted value of the dependent variable y_j.

Given that $\underset{\sim}{X}$ is a matrix of order $n \times p$ of X values (deviation scores $x_{ji} - \bar{x}_i$), the vector of regression weights is estimated by

$$\underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}' \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{Y}, \qquad [3]$$

where $\underset{\sim}{Y}$ is a column vector of elements y_j and $\underset{\sim}{X}'$ is the transpose of $\underset{\sim}{X}$. The error variance σ_e^2 (where $e_j = y_j - \hat{y}_j$) is estimated by

$$s_e^2 = \Sigma_j e_j^2 / (n - 1), \quad [4]$$

and the estimates of the error variances of the $\hat{\beta}$'s are given by the respective diagonal elements of the covariance matrix

$$\text{cov}(\hat{\beta}) = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} s_e^2. \quad [5]$$

The above equations assume that the independent variables are measured without error. To the extent that this is not true, the estimates of their variances will be inflated. That is, the diagonal elements of the matrix $\underset{\sim}{X}'\underset{\sim}{X}$ will be larger than they should otherwise be. In addition, since the x's are random variables chosen as plausible predictors of y, it is possible (even probable) that the estimate of error variance s_e^2 (Equation 4) will be an overestimate of the true error variance of the y_j's.

The first of these effects comes into play when estimating the values of the regression coefficients in Equation 3. Because that equation involves the inverse of the matrix $\underset{\sim}{X}'\underset{\sim}{X}$, the regression coefficients are necessarily underestimated. Both of the effects mentioned above play a part in the estimation of the covariance matrix in Equation 5. There can never be certainty that these effects will cancel out each other. Maxwell (1975) cautions:

> In summary we see that inadequate specification of y and errors of measurement in the x's lead to a situation in which the tests of significance provided for the classic model are of dubious validity in most social science applications. At best we can claim that, if e_j are calculated and found to be approximately normally distributed, a significant multiple correlation coefficient would indicate some dependence of y on a weighted sum of the x's. But the relative sizes of the regression weights would be suspect and the magnitude of the multiple correlation coefficient in particular would be the point to note. (pp. 52-53)

Both Lawley and Maxwell (1973) and Maxwell (1975) show how such errors of measurement in the x's can be handled by stating the model in factor analytic terms and proceeding from there. Essentially, the set of predictor variables is reduced to a "best" set of statistically independent variables (i.e., the factors), and then the dependent variable is predicted from these. Specifically, the analysis proceeded as follows:

The maximum likelihood estimate of the correlation matrix is given by

$$\underset{\sim}{\Sigma}^* = \underset{\sim}{\Lambda}^* \underset{\sim}{\Lambda}^{*\prime} + \underset{\sim}{\Psi}^*, \quad [6]$$

where

$\underset{\sim}{\Sigma}^*$ (of order $1 + p$) includes the dependent variable y together with the p independent variables,

$\underset{\sim}{\Lambda}^*$ is a $(1 + p) \times k$ matrix of factor loadings of all the variables on the k factors, and

$\underset{\sim}{\Psi}^*$ is a diagonal matrix of residual variances.

Partitioning $\underset{\sim}{\Lambda}^*$ as

$$\underset{\sim}{\Lambda}^* = \begin{bmatrix} \underset{\sim}{\lambda}_1' \\ \underset{\sim}{\Lambda} \end{bmatrix}, \qquad [7]$$

where $\underset{\sim}{\lambda}_1'$ contains the loadings of y on the factors and $\underset{\sim}{\Lambda}$ contains the corresponding loadings of the x's, yields the regression equation

$$\hat{\underset{\sim}{y}} = \underset{\sim}{\lambda}_1' \underset{\sim}{f} . \qquad [8]$$

Estimating the factors $\underset{\sim}{f}$ in this equation (see Maxwell, 1975, p. 59) yields the new regression equation

$$\hat{\underset{\sim}{y}} = \underset{\sim}{\lambda}_1' \underset{\sim}{\Gamma}^{-1} \underset{\sim}{\Lambda}' \underset{\sim}{\Psi}^{-1} \underset{\sim}{x} , \qquad [9]$$

where $\underset{\sim}{\Gamma} = \underset{\sim}{\Lambda}' \underset{\sim}{\Psi}^{-1} \underset{\sim}{\Lambda}$ is a diagonal matrix. In this approach, the square of the multiple correlation coefficient for the y's predicted from the x's is given by the communality of y in the maximum likelihood factor analysis.

For this study, maximum likelihood factor analyses were performed separately on the 3×3, 4×4, and 5×5 Σ^* matrices corresponding to the 2, 3, and 4 independent variable cases, respectively (the dependent variable y is always included in the Σ^* matrix). The matrices from a one-factor solution were obtained in each case and Equation 9 was calculated for predicting scores on Subtests 3, 4, and 5, respectively, from the scores on all previously administered subtests.

To examine the effect of using the corrected (versus the classical) regression equations, the subtests were administered in the same order for inter-subtest branching as they were for the classical equations. Since factor analyses cannot be performed when the number of variables is less than three, the classical regression equations were used for the prediction of Subtest 2 scores.

Since the square of the multiple correlation coefficient (R) was given by the communality of y in these analyses, the standard error of estimate (SEE) was computed using the formula

$$SEE = s_y \sqrt{1 - R^2} \qquad [10]$$

Cross-validation. Since this study was a real-data simulation of various testing strategies, the regression equations developed from students' subtest scores during any one academic quarter were used in the inter-subtest branching strategy simulated from students' item responses from that same quarter. As with any application of multiple regression techniques, the estimates of the

b-weights and the multiple correlation coefficient were likely to be inflated due to sample-specificity. To the extent that this was true, the inter-subtest branching strategy would be nonoptimal for any subsequent sample of students.

To investigate the extent to which variance in the multiple correlation coefficients and the *b*-weights affected the efficacy of the inter-subtest branching strategy employed here, a double-cross-validation design was used. Both the fall and winter quarter samples served as independent development groups, and both sets of regression equations (classical and corrected) were obtained separately for each group. Then, the equations developed from the fall data were used in the simulation with the data from both the fall and winter quarters and correspondingly for the equations developed from the winter data. The results obtained in this way allowed for a direct investigation of the extent to which the efficacy of the adaptive strategies was affected by cross-sample discrepancies in the regression equations.

Adaptive Intra-Subtest Item Selection

Brown and Weiss (1977) compared the results obtained from the entire testing strategy combining both intra-subtest item selection and inter-subtest branching with those obtained when the tests were conventionally administered. In this study the effects of the variable termination criterion in the intra-subtest item selection strategy were separated from those of the inter-subtest branching strategy, and the relative contributions of these aspects of the adaptive strategy were determined.

Consequently, a third set of testing conditions was simulated. Here, the five subtests were treated as independent sets of items. Instead of branching from one subtest to the next using the regression-based inter-subtest branching strategy, each subtest was considered to be a self-contained test. As in the conventional test, Bayesian scoring was used; and a mean of 0.0 with a variance of 1.0 was used as the initial prior $\hat{\theta}$ for each of the five subtests. Items within each subtest, however, were selected according to the intra-subtest item selection scheme described above, and the variable termination information criterion values of .01 and .05 were used. Hence, the only difference between these tests and the other sets of adaptive tests was that inter-subtest branching was not utilized here.

Dependent Variables

The important question in this study was not "Can test length be reduced by adaptive testing?" but rather "Can test length be reduced and adequate levels of measurement precision be maintained?" It would be pointless to reduce test length by 20%, 30%, or more if much of the measurement accuracy was sacrificed in the process.

Correlations of Achievement Level Estimates

One means of investigating the extent to which measurement precision was preserved or lost by the adaptive testing strategy is correlational analysis; that is, how well did the achievement estimates on the adaptive tests correlate with those on the conventional tests? For this study these correlations were obtained for each of the subtests across all testing conditions.

Information

The degree to which measurement precision is lost through test-length reduction may also be assessed by inspection of the relevant subtest information curves. The adaptive subtest information curves were obtained as follows:

A student's final $\hat{\theta}$ was obtained for any one subtest after testing terminated for that subtest. Then, the item information function (Birnbaum, 1968) was evaluated at that student's final $\hat{\theta}$ for each item that was administered adaptively. These item information values were then summed across all items administered to the student in that subtest in order to obtain the adaptive subtest information curve for that student.

The conventional subtest information curves were obtained in essentially the same way, except that the item information functions were evaluated at the $\hat{\theta}$ arising from administration of the conventional subtest, and they were summed over all the items in the subtest pool.

When a final $\hat{\theta}$ had been obtained for every student, the students were grouped into 20 nonoverlapping intervals on the basis of their $\hat{\theta}$ values from either the conventional or adaptive test. The mean subtest information value (over all students within an interval) was obtained for each of the 20 intervals separately for the conventional and adaptive tests; these mean values were then plotted at the midpoint of each interval in order to obtain the subtest information curves.

RESULTS

Preliminary Results

Item Parameters

Table 1 presents the means and standard deviations for estimates of the latent trait item parameters a, b, and c. Also included are the number and percentage of items from the final exams for which parameter estimates could be obtained. Individual item parameter estimates, by subtest, are shown in Appendix Tables A and B for the fall and winter data, respectively.

Table 1 shows that item parameters were obtained for 94% (or 46) of the 49 items available on the fall quarter final exam. This retention rate ranged from 85% of the items in the Chemistry subtest to 100% of the items in the Cell, Energy, and Reproduction subtests. The winter quarter final exam exhibited a somewhat lower retention rate, with 84% (or 31) of the 37 available items yielding parameter estimates. The Ecology subtest suffered the largest loss (75% retention), although closer inspection revealed that this was a loss of only 1 of the 4 original items; no subtest lost more than 2 items. In terms of absolute numbers of items, the winter quarter item pool was somewhat smaller than that from fall quarter: 31 parameterized items compared to 46.

The overall mean b parameter for the fall quarter item pool (-.22) was slightly lower than that for the winter quarter pool, $\bar{b}$ =.02. The mean a parameters of 1.80 and 1.81 and c parameter of .40 were essentially identical for the two pools.

Table 1
Means and Standard Deviations of Normal Ogive Item Discrimination (a), Difficulty (b), and Lower Asymptote (c) Parameter Estimates for the Fall and Winter Quarter Final Exams by Subtest

Quarter and Subtest	Number of Items Available	Number of Items Parameterized	Percent of Items Parameterized	a Mean	a *SD*	b Mean	b *SD*	c Mean	c *SD*
Fall									
Chemistry	13	11	85	1.56	.44	-.49	.78	.32	.09
Cell	9	9	100	1.84	.41	.23	1.34	.45	.09
Energy	9	9	100	2.27	.47	-.05	1.02	.42	.13
Reproduction	11	11	100	1.64	.57	-.13	.92	.40	.14
Ecology	7	6	86	1.73	.36	-.80	.67	.44	.07
Total	49	46	94	1.80	.51	-.22	.99	.40	.12
Winter									
Chemistry	10	8	80	1.77	.37	-.29	.82	.29	.07
Cell	6	6	100	1.69	.26	-.09	1.06	.38	.07
Energy	8	7	88	2.22	.49	.21	.79	.45	.14
Reproduction	9	7	78	1.53	.32	.25	1.22	.47	.11
Ecology	4	3	75	1.81	.54	.08	1.64	.51	.24
Total	37	31	84	1.81	.44	.02	1.00	.40	.14

Ordering of Subtests

The intercorrelations of Bayesian ability estimates from the five subtests in each quarter are shown in Table 2. For the data from fall quarter, these inter-subtest correlations ranged from .289 (between Ecology and Energy) to .433 (between Cell and Chemistry). The range of correlations was somewhat larger for the winter quarter data; the lowest correlation was .160 (between Cell and Ecology) and the largest correlation was .496 (between Chemistry and Energy).

Since the highest correlation was between Chemistry and Cell in the fall data and between Chemistry and Energy in the winter data, the Chemistry subtest was designated to be administered first in each case; the Cell subtest was administered second for the fall quarter equations and the Energy subtest was administered second for the winter quarter equations.

Table 2
Intercorrelations of Bayesian Ability Estimates on the Five Subtests of the Fall (Below Diagonal) and Winter (Above Diagonal) Quarter Final Exams

Subtest	Chemistry	Cell	Energy	Reproduction	Ecology
Chemistry		.451	.496	.379	.228
Cell	.433		.456	.301	.160
Energy	.412	.370		.347	.189
Reproduction	.388	.344	.321		.221
Ecology	.387	.302	.289	.302	

For the fall quarter data, multiple regression equations were obtained using the Chemistry and Cell subtests as independent variables and each of the other subtests, in turn, as the dependent variable. Because the Energy subtest had the highest multiple correlation with these first two subtests, it was chosen as the third subtest to be administered. This procedure was repeated to select the fourth and fifth subtests for administration. The same process was carried out using the winter quarter data.

Appendix Table C shows the intermediate classical regression equations used to choose the order of administration of the subtests for both fall and winter quarters. For the fall equations the subtests were ordered in the following sequence: Chemistry, Cell, Energy, Reproduction, and Ecology. For the winter equations the order was Chemistry, Energy, Cell, Reproduction, and Ecology.

Table 3 shows the classical (or uncorrected) regression coefficients, multiple correlation coefficients, and standard errors of estimate for the sets of regression equations from both the fall and winter data. These equations were those used for inter-subtest branching.

Table 3
Regression Coefficients, Multiple Correlation Coefficients (R), and Standard Errors of Estimate (SEE) for the Classical Regression Equations from the Fall and Winter Quarter Final Exams

Quarter and Criterion Subtest	Regression Coefficients for Scores on Previously Administered Subtests				Regression Constant	R	SEE
	Chemistry	Cell	Energy	Reproduction			
Fall							
Cell	.400				.137	.433	.680
Energy	.328	.272			-.009	.464	.768
Reproduction	.240	.190	.140		.204	.455	.707
Ecology	.221	.110	.089	.128	-.029	.446	.665
Winter							
Energy	.461				.056	.496	.637
Cell	.276		.305		-.144	.525	.620
Reproduction	.258	.129	.203		.134	.432	.761
Ecology	.102	.026	.052	.103	.112	.278	.595

Corrected Equations

The corrected regression coefficients, multiple correlation coefficients, and standard errors of estimate from the fall and winter final exams are given in Table 4. The factor loadings and estimates of communalities used to compute these equations are given in Appendix Table D. It should be noted that the factor analytic techniques could not be applied, of course, unless there were at least three variables in the regression equation. Hence, for the cases in which there were only two variables, e.g., one predictor subtest and one criterion subtest, the classical (or uncorrected) regression equation was used. Therefore, the first and fifth lines in Table 4 match exactly the first and fifth lines, respectively, of Table 3.

Table 4
Regression Coefficients, Multiple Correlation Coefficients (R), and Standard Errors of Estimate (SEE) for the Corrected Regression Equations from the Fall and Winter Quarter Final Exams

Quarter and Criterion Subtest	Regression Coefficients for Scores on Previously Administered Subtests				Regression Constant	R	SEE
	Chemistry	Cell	Energy	Reproduction			
Fall							
Cell	.400				.137	.433	.680
Energy	.538	.446			-.008	.594	.698
Reproduction	.345	.279	.216		.206	.552	.662
Ecology	.266	.195	.152	.152	-.024	.523	.633
Winter							
Energy	.461				.056	.496	.637
Cell	.416		.461		-.132	.644	.557
Reproduction	.296	.230	.295		.153	.504	.729
Ecology	.119	.088	.113	.051	.127	.303	.590

Comparison of the entries in Table 3 with those in Table 4 reveals that the Lawley-Maxwell method of correction for multiple regression equations did indeed increase the sizes of both the multiple correlation coefficient and the regression coefficients. Inspection of the fall quarter data, for example, shows that the corrected multiple correlation coefficients increased from R = .464, .455, and .446 to R = .594, .552, and .523, respectively; there were corresponding decreases in the sizes of the standard errors of estimate. The b-weights also increased in size, with the largest increases occurring in those equations with the fewest independent variables. For example, when the Energy subtest was the criterion, the regresssion coefficients for the Chemistry and Cell subtests increased from b = .328 and .272 to b = .538 and .446, respectively.

A similar effect was observed with the winter quarter data. Here, the corrected multiple correlation coefficients increased from R = .525, .432, and .278 to R = .644, .504, and .303, respectively; again, there were corresponding decreases in the sizes of the standard errors of estimate. All but one of the b-weights increased in size; the b-weight for the Reproduction subtest in the final equation decreased from .103 to .051.

Test Length

Mean Test Length

Table 5 presents the mean numbers of items administered in each of the five subtests and in the total test for the conventional test and for the adaptive test using adaptive intra-subtest item selection but no inter-subtest branching.

Conventional test. During the actual final exam in each quarter, students were free to omit any 10 (of 110) items of their choice. To the extent that students omitted some of the items with ICC parameters that were selected for inclusion in these simulation item pools (i.e., from the five content areas--

Chemistry, Cell, Energy, Reproduction, and Ecology), the number of items for which student responses were available varied across students. Thus, in these five content areas, students answered from 37 to 46 of the parameterized items in fall and 23 to 31 items in winter. Consequently, the conventionally administered test was, on the average, 43 items long for the fall quarter data and 28.55 items long for the winter data.

Table 5
Number of Items Administered in the Five Subtests of the Fall and Winter Quarter Final Exams with No Inter-Subtest Branching

					Adaptive Intra-Subtest Item Selection: Termination Criterion							
	Conventional Test				.01				.05			
Subtest and Data	Mean	*SD*	Range Min	Max	Mean	*SD*	Range Min	Max	Mean	*SD*	Range Min	Max
Chemistry												
Fall	10.21	.91	6	11	9.13	1.41	5	11	8.09	1.59	4	11
Winter	7.48	.72	4	8	6.59	1.16	3	8	5.85	1.16	2	8
Cell												
Fall	8.50	.71	5	9	6.93	.89	3	8	5.68	1.10	3	7
Winter	5.64	.60	3	6	4.73	.85	2	6	4.26	.71	2	5
Energy												
Fall	8.09	.95	4	9	5.96	1.03	3	9	5.15	.88	2	8
Winter	5.91	1.01	2	7	4.67	.95	2	7	4.30	1.03	2	7
Reproduction												
Fall	10.46	.84	7	11	8.78	1.08	4	11	7.67	1.33	4	10
Winter	6.69	.56	3	7	4.93	1.09	1	7	4.04	.80	1	5
Ecology												
Fall	5.73	.50	3	6	5.24	.74	2	6	4.07	1.20	2	6
Winter	2.82	.38	2	3	1.95	.21	1	2	1.07	.26	1	2
Total Test												
Fall	43.00	1.77	37	46	36.04	2.46	28	42	30.67	3.17	22	41
Winter	28.55	1.60	23	31	22.87	2.47	14	29	19.52	2.12	12	26

The discrepancy between the two quarters in the numbers of items available in the conventional test for this study was fairly evenly distributed across all five subtests, so that the relative size of each subtest remained about the same (see Table 1). That is, Chemistry and Reproduction were the longest subtests, and Ecology was consistently the shortest.

Adaptive intra-subtest item selection. In these sets of tests, the intra-subtest item selection strategy was employed with a variable termination criterion, but no inter-subtest branching scheme was used. That is, a prior $\hat{\theta}$ of 0.0 with an estimated variance of 1.0 was used as an entry point in each of the five subtests. Table 5 shows data on test lengths obtained for each subtest under the two termination criteria used in this study (item information of .01 and .05). During the fall quarter the length of the total test battery averaged 36.04 items under the more stringent termination criterion, .01, and 30.67 items under the termination criterion of .05. For winter quarter these figures were 22.87 and 19.52, respectively.

In all cases the maximum number of items administered under this adaptive strategy represented some reduction in total test battery length. For the fall data no student answered more than 42 items under the .01 termination criterion; and the shortest adaptive test was only 28 items long. For the .05 criterion the longest test was 41 items; the shortest was 22. For the winter quarter data these figures were 29 and 14 for the .01 termination criterion and 26 and 12 for the .05 criterion.

Inter-subtest branching. When the inter-subtest branching strategy was employed in addition to the adaptive intra-subtest item selection strategy and variable termination criterion, test length was reduced even further. Tables 6 and 7 show the mean test lengths under these conditions, when both the classical and corrected regression equations were developed on the data from the fall and winter quarters, respectively. Data for the Chemistry subtest (the first subtest administered) are the same in the two tables because the initial $\hat{\theta}$ was assumed to be 0.0 with a variance of 1.0 for all students and was constant for the first subtest, regardless of branching strategy used (e.g., no branching versus inter-subtest branching).

For both the .01 and .05 termination criterion, the addition of the inter-subtest branching strategy generally resulted in shorter tests; the exception was the Ecology subtest with a .05 termination criterion under all testing conditions. However, in comparison to the results from use of intra-subtest branching only (see Table 5), this reduction was slight--never more than one item for the total test. The data also show that the branching strategy utilizing the corrected regression equations resulted in tests that were shorter than when the classical regression equations were used, although the difference was very slight. For example, under the .01 termination criterion, the classical fall quarter regression equations resulted in a total test battery length of 35.61 items for the fall data and 35.15 items when the corrected regression equations were used (Table 6). When the .05 termination criterion was used, the classical fall quarter equations resulted in a mean test battery length of 30.33 items versus 30.10 items for the corrected equations. There was a tendency for the corrected equations to result in higher standard deviations of numbers of items administered in the total test than did the classical equations; this was due to the tendency toward shorter minimum total test lengths. Similar results were observed when the winter quarter equations were used (see Table 7).

Cross-validation. There was very little difference between total test lengths in the development groups and in cross-validation; the differences which were found were usually in the direction of shorter tests when the regression equations were cross-validated on data from the other quarter. For example, when the classical regression equations developed on winter quarter data were applied to that same data, mean test length was 22.64 and 19.90 for termination criteria of .01 and .05, respectively (see Table 7). When the cross-validated classical fall quarter equations were applied to that winter data (Table 6), however, the means were 22.58 and 19.68, respectively. The results for the classical regression equations applied to the fall quarter data were mixed. When the results from the sets of corrected equations were compared, they favored the cross-validated condition whenever a difference was found.

Table 6

Number of Items Administered in the Five Subtests of the Fall and Winter Quarter Final Exams for the Adaptive Test with Intra-Subtest Item Selection and Inter-Subtest Branching Using Classical and Corrected Regression Equations from Fall Data

	Classical Equations: Termination Criterion								Corrected Equations: Termination Criterion							
	.01				.05				.01				.05			
			Range				Range				Range				Range	
Subtest and Data	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Chemistry																
Fall	9.13	1.41	5	11	8.09	1.59	4	11	9.13	1.41	5	11	8.09	1.59	4	11
Winter	6.59	1.16	3	8	5.85	1.16	2	8	6.59	1.16	3	8	5.85	1.16	2	8
Cell																
Fall	6.78	.84	4	8	5.54	1.34	2	8	6.78	.84	4	8	5.54	1.34	2	8
Winter	4.64	.79	2	6	4.07	.89	1	5	4.64	.79	2	6	4.07	.89	1	5
Energy																
Fall	5.84	1.20	2	9	4.91	1.11	2	8	5.66	1.33	2	9	4.77	1.27	1	8
Winter	4.57	1.21	1	7	4.14	1.30	1	7	4.39	1.37	1	7	3.92	1.48	0	7
Reproduction																
Fall	8.67	1.17	5	11	7.58	1.41	3	10	8.51	1.34	4	11	7.50	1.55	2	10
Winter	4.92	.97	1	7	4.13	.88	1	6	4.83	1.06	1	7	4.06	.93	1	7
Ecology																
Fall	5.19	.79	2	6	4.22	1.25	1	6	5.06	.94	2	6	4.20	1.28	1	6
Winter	1.86	.35	0	2	1.50	.51	0	2	1.78	.42	0	2	1.50	.51	0	2
Total Test																
Fall	35.61	2.94	24	43	30.33	3.81	18	41	35.15	3.44	22	43	30.10	4.16	15	41
Winter	22.58	2.87	13	29	19.68	2.64	11	26	22.24	3.12	12	29	19.40	2.86	10	26

Note. Winter data is cross-validation.

Table 7

Number of Items Administered in the Five Subtests of the Fall and Winter Quarter Final Exams for the Adaptive Test with Intra-Subtest Item Selection and Inter-Subtest Branching Using Classical and Corrected Regression Equations from Winter Data

	Classical Equations: Termination Criterion								Corrected Equations: Termination Criterion							
	.01				.05				.01				.05			
Subtest and Data	Mean	*SD*	Range Min	Range Max	Mean	*SD*	Range Min	Range Max	Mean	*SD*	Range Min	Range Max	Mean	*SD*	Range Min	Range Max
Chemistry																
Winter	6.59	1.16	3	8	5.85	1.16	2	8	6.59	1.16	3	ε	5.85	1.16	2	8
Fall	9.13	1.41	5	11	8.09	1.59	4	11	9.13	1.41	5	11	8.09	1.59	4	11
Energy																
Winter	4.69	1.16	1	7	4.28	1.19	1	7	4.69	1.16	1	7	4.28	1.19	1	7
Fall	5.92	1.17	2	9	5.05	1.02	2	8	5.92	1.17	2	9	5.05	1.02	2	8
Cell																
Winter	4.54	.80	2	6	4.02	.84	2	5	4.50	.82	2	6	3.93	.93	1	5
Fall	6.62	1.00	2	8	5.59	1.34	2	8	6.33	1.24	2	8	5.29	1.60	1	8
Reproduction																
Winter	4.86	1.03	1	7	4.09	.88	1	6	4.79	1.07	1	7	4.01	.90	1	7
Fall	8.66	1.20	4	11	7.59	1.43	2	10	8.53	1.39	4	11	7.51	1.58	2	10
Ecology																
Winter	1.95	.21	1	2	1.68	.47	0	2	1.87	.34	0	2	1.34	.48	0	2
Fall	5.23	.76	2	6	4.41	1.09	2	6	5.24	.74	2	6	4.15	1.29	1	6
Total Test																
Winter	22.64	2.87	13	29	19.90	2.68	11	26	22.44	2.98	13	29	19.40	2.54	10	25
Fall	35.56	2.95	22	43	30.73	3.79	17	40	35.14	3.45	21	43	30.09	4.16	16	40

Note. The results from the winter data are presented before those from fall in this table because the winter data represent the development group, and the fall data the cross-validation group.

Percent Reduction in Test Length

Table 8 summarizes the percent reduction in the mean number of items administered in each subtest and in the total test under the various testing conditions.

Adaptive intra-subtest item selection. The first column of data in Table 8 represents the reduction in mean test length that was observed when only the adaptive intra-subtest item selection strategy with a variable termination criterion was compared to a conventionally administered test. In both these adaptive and conventional tests, each subtest was treated as a separate unit with no inter-subtest branching between tests. For the fall quarter data, use of the adaptive testing strategy decreased total test length by 16.19% under the .01 termination criterion and decreased it by as much as 28.67% when the .05 criterion was used. When this strategy was used on the winter quarter data, the respective reductions were 19.89% and 31.63% in total test length.

The largest reduction in subtest length using a termination criterion of .01 occurred for the fifth subtest, Ecology, and amounted to a total decrease of almost 31% of the items. This effect, however, was limited to the winter data, as the Ecology subtest for the fall data exhibited a reduction of less than 9%. On the average, the Chemistry subtest (the first subtest administered) showed the smallest decrease in number of items administered--about 10 to 12%. The same pattern was observed among the subtests when a termination criterion of .05 was used. That is, the largest reduction in subtest length was observed for the Ecology subtest for the winter data (62.06%); and the smallest reduction, on the Chemistry subtest for the fall data (20.76%).

Inter-subtest branching. The remaining columns of Table 8 show the results obtained when the inter-subtest branching scheme was coupled with the adaptive intra-subtest item selection strategy and then compared to a conventionally administered test. The reductions in total test length were slightly greater than those obtained when the inter-subtest branching strategy was not utilized.

For example, when the fall quarter equations were applied to the fall quarter data, the reduction in average test length for the total test increased from 16.19% to 17.19% for the classical equations and 18.26% for the corrected equations under the .01 termination criterion. These figures were 28.67%, 29.47%, and 30.00%, respectively, for the .05 termination criterion. Use of the corrected regression equations generally resulted in somewhat shorter total test lengths than did use of the classical equations, although the difference was slight.

When the winter quarter equations were applied to the winter quarter data, total test length was reduced from 19.89% to 20.70% for the classical equations and 21.40% for the corrected equations under the .01 termination criterion. These figures were 31.63%, 30.30%, and 32.05%, respectively, for the .05 termination criterion. Use of the classical equations actually resulted in tests which were slightly longer under the .05 criterion than when no inter-subtest branching strategy was used. Use of the corrected equations, however, resulted in shorter tests, as expected.

In general (across both sets of data), additional reduction in test length was less than three percentage points, and most often one percentage point or

Table 8

Percent Reduction from the Conventional Test in Mean Number of Items Administered in the Five Subtests of the Fall and Winter Quarter Final Exams With and Without Inter-Subtest Branching Using Classical and Corrected Regression Equations Developed from Each Quarter

Subtest and Data	Adaptive Intra Subtest Item Selection: Termination Criterion		Percent Mean Reduction[a] Due to Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching							
			Classical Equations				Corrected Equations			
			Fall: Termination Criterion		Winter: Termination Criterion		Fall: Termination Criterion		Winter: Termination Criterion	
	.01	.05	.01	.05	.01	.05	.01	.05	.01	.05
Chemistry										
Fall	10.58	20.76	10.58	20.76	10.58	20.76	10.58	20.76	10.58	20.76
Winter	11.90	21.79	11.90	21.79	11.90	21.79	11.90	21.79	11.90	21.79
Cell										
Fall	18.47	33.18	20.24	34.82	22.12	34.24	20.24	34.82	25.53	37.76
Winter	16.13	24.47	17.73	27.84	19.50	28.72	17.73	27.84	20.21	30.32
Energy										
Fall	26.33	36.34	27.81	39.31	26.82	37.58	30.04	41.04	26.82	37.58
Winter	20.98	27.24	22.67	29.95	20.64	27.58	25.72	33.67	20.64	27.58
Reproduction										
Fall	16.06	26.67	17.11	27.53	17.21	27.44	18.64	28.30	18.45	28.20
Winter	26.31	39.61	26.46	38.27	27.35	38.86	27.80	39.31	28.40	40.06
Ecology										
Fall	8.55	28.97	9.42	26.35	8.73	23.04	11.69	26.70	8.55	27.57
Winter	30.85	62.06	34.04	46.81	30.85	40.43	36.88	46.81	33.69	52.48
Total Test										
Fall	16.19	28.67	17.19	29.47	17.30	28.53	18.26	30.00	18.28	30.02
Winter	19.89	31.63	20.91	31.07	20.70	30.30	22.10	32.05	21.40	32.05

[a]Computed by the formula: 100-[(Mean number of items in appropriate adaptive test/mean number of items in conventional test) × 100].

less. Use of the corrected equations resulted in shorter tests in all cases in comparison with use of adaptive intra-subtest item selection alone. The Energy subtest showed the largest decreases in test length across testing conditions (with the exception of the Ecology subtest administered during winter quarter, which showed the greatest reduction in test length). This was followed closely by the Cell, Reproduction, and Chemistry subtests, respectively. During fall quarter the decrease in the length of the Ecology subtest was the smallest.

Cross-validation. When the fall quarter equations were applied to the data from winter quarter in the cross-validation condition, test-length reduction increased from 19.89% with no inter-subtest branching to 20.91% for the classical equations and 22.10% for the corrected equations, under the .01 termination criterion. For the termination criterion of .05,these figures were 31.63% with no inter-subtest branching and 31.07% and 32.05% for the two inter-subtest branching conditions with .01 and .05 termination, respectively. With the winter data there was a slight increase in test length on cross-validation from 28.67% without inter-subtest branching to 30.30% for the classical equations and .05 termination criterion.

For the double-cross-validation condition, when the winter quarter equations were applied to the fall quarter data, reductions in test length were again observed. For the .01 termination criterion, test length decreased from 16.19% without inter-subtest branching to 17.30% for the classical equations and 18.28% for the corrected equations. These figures were 28.67%, 28.53%, and 30.02%, respectively, for the .05 termination criterion. (Only with the .01 termination criterion were the tests with the cross-validated equations consistently shorter than the tests with the original (development group) equations. At the .05 termination level the results from the classical and corrected equations were mixed.

In summary, for the .01 termination criterion the reduction in total test length for the data from each of the quarters was nearly always greater when the regression equations were cross-validated. The results from using the .05 criterion were mixed. As was observed with the two development groups, use of the corrected equations resulted in shorter mean test lengths under cross-validation than did use of the cross-validated classical equations. In all cases, however, observed differences in test length reduction were slight.

Minimum and maximum reductions in test length. The data in Table 8 reflect only the reductions in average test lengths. Table 9 presents the minimum and maximum reductions from the conventional test length that were observed for any one student when the inter-subtest branching strategy was used. Inspection of this table reveals that for each testing condition (except for the corrected fall equations applied to the winter data with .01 termination criterion), total test length was reduced for all students by at least 2.5%. The largest reduction in total test length was that observed for the fall data using corrected fall equations and a termination criterion of .05, where the reduction was 67.4%.

For each subtest separately the minimum reduction in subtest length (for all tests but one) was 0%; that is, there was at least one student who was administered all the available items in a subtest regardless of testing condition. However, there also were students whose subtests were reduced in length by more

Table 9

Minimum and Maximum Percent Reduction from the Conventional Test Length Observed for Any One Student When the Adaptive Inter-Subtest Branching Strategy Was Used in the Five Subtests of the Fall and Winter Quarter Final Exams

	Classical Equations								Corrected Equations							
	Fall: Termination Criterion				Winter: Termination Criterion				Fall: Termination Criterion				Winter: Termination Criterion			
Subtest	.01		.05		.01		.05		.01		.05		.01		.05	
and Data	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Chemistry																
Fall	0.0	37.5	0.0	50.0	0.0	37.5	0.0	50.0	0.0	37.5	0.0	50.0	0.0	37.5	0.0	50.0
Winter	0.0	42.9	0.0	57.1	0.0	42.9	0.0	57.1	0.0	42.9	0.0	57.1	0.0	42.9	0.0	57.1
Cell																
Fall	0.0	55.6	0.0	75.0	0.0	66.7	0.0	77.8	0.0	55.6	0.0	75.0	0.0	77.8	0.0	88.9
Winter	0.0	50.0	0.0	75.0	0.0	50.0	0.0	66.7	0.0	50.0	0.0	75.0	0.0	66.7	0.0	83.3
Energy																
Fall	0.0	77.8	0.0	77.8	0.0	77.8	0.0	77.8	0.0	77.8	0.0	88.9	0.0	77.8	0.0	77.8
Winter	0.0	71.4	0.0	85.7	0.0	66.7	0.0	85.7	0.0	85.7	0.0	100.0	0.0	60.0	0.0	85.7
Reproduction																
Fall	0.0	54.5	0.0	72.7	0.0	54.5	0.0	81.8	0.0	54.5	0.0	81.8	0.0	54.5	0.0	81.8
Winter	0.0	80.0	0.0	80.0	0.0	80.0	14.3	80.0	0.0	80.0	0.0	80.0	0.0	80.0	0.0	80.0
Ecology																
Fall	0.0	50.0	0.0	75.0	0.0	40.0	0.0	60.0	0.0	60.0	0.0	75.0	0.0	40.0	0.0	75.0
Winter	0.0	100.0	0.0	100.0	0.0	50.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0
Total Test																
Fall	2.5	45.7	9.8	60.9	2.5	52.2	7.5	63.0	2.5	47.8	5.0	67.4	2.5	54.3	7.5	65.2
Winter	3.3	50.0	12.0	58.6	3.3	50.0	12.0	58.6	0.0	53.8	13.3	65.5	3.3	50.0	14.3	60.7

than 75%. In fact, there were some subtests (specifically, Ecology) that students "skipped" altogether, as evidenced by the 100% maximum reduction figures for most of the winter data.

It would be expected that as the tests continued and more information was available with which to predict scores on subsequent subtests, these predicted scores--hence, entry points into the subtest--would become more accurate. This should be reflected in more stable ability estimates and therefore shorter subsequent subtests. Indeed, there is a trend in the data of Table 9 for increasingly larger reductions in test length for the tests administered later in the inter-subtest branching.

Correlations of Achievement Level Estimates

Table 10 presents the values of the correlation coefficients (r) between the Bayesian $\hat{\theta}$ values from the conventional tests and the adaptive tests, under all testing conditions. Generally, these correlations were fairly homogeneous; more than half of them were greater than .90, while less than 10% of them were below .80.

Adaptive Intra-Subtest Item Selection

With no inter-subtest branching, the largest correlations were those observed for the Cell subtest with variable termination .01--for both sets of data, r = .998; and for the Ecology subtest under the same conditions for winter data, r = .995. The smallest correlation was observed for the Ecology subtest with a termination criterion of .05; here, the winter data correlation was r = .527. This appears rather low, but the average length of this adapted subtest was only 1.07 items (see Table 5).

Inter-Subtest Branching

Classical equations. When the classical fall quarter equations were applied to the data collected from that same quarter, the range of correlations was fairly small. These correlations ranged from .846 (for the Energy subtest) to .979 (for the Cell subtest) with the .01 termination criterion. For the termination criterion of .05, these correlations were .795 (for Energy) and .890 (for both Reproduction and Ecology).

When the winter quarter equations were applied to the winter data, the correlations varied even less. For the .01 termination criterion the range was from .921(for Reproduction) to .983 (for Chemistry). For the .05 criterion the range was from .876 (for Reproduction) to .962 (for Chemistry).

In general, the addition of an inter-subtest branching strategy to adaptive intra-subtest item selection reduced the correlations between conventional and adaptive subtest scores by a small amount (less than .021 for the fall data and less than .040 for the winter data). The single exception to this was for the winter administration of the Ecology subtest (termination criterion of .05), where inter-subtest branching increased the correlation from .527 to .886. These reductions in the correlations can be accounted for by the decreases in number of items with which θ was estimated; the inter-subtest branching strategy typically reduced test length over that obtained with intra-subtest

Table 10

Correlations of Bayesian Achievement Level Estimates for the Adaptive and Conventional Testing Strategies for Each Subtest of the Fall and Winter Quarter Final Exams

Subtest and Data	Adaptive Intra-Subtest Item Selection: Termination Criterion		Adaptive Inter-Subtest Item Selection with Intra-Subtest Branching							
			Classical Equations				Corrected Equations			
			Fall: Termination Criterion		Winter: Termination Criterion		Fall: Termination Criterion		Winter: Termination Criterion	
	.01	.05	.01	.05	.01	.05	.01	.05	.01	.05
Chemistry										
Fall	.941	.887	.941	.887	.941	.887	.941	.887	.941	.887
Winter	.983	.962	.983	.962	.983	.962	.983	.962	.983	.962
Cell										
Fall	.998	.873	.979	.858	.966	.883	.979	.858	.889	.830
Winter	.998	.964	.972	.924	.960	.935	.972	.924	.918	.879
Energy										
Fall	.852	.808	.846	.795	.842	.792	.818	.770	.842	.792
Winter	.989	.943	.972	.914	.967	.923	.926	.882	.967	.923
Reproduction										
Fall	.942	.909	.924	.890	.926	.891	.904	.871	.914	.873
Winter	.941	.898	.926	.862	.921	.876	.895	.833	.889	.836
Ecology										
Fall	.940	.871	.919	.890	.936	.912	.889	.863	.928	.882
Winter	.995	.527	.887	.759	.958	.886	.768	.671	.917	.715

item selection alone. This effect can also be seen by comparing the results from the two termination criteria; the correlations were typically lower for the .05 criterion, which generally yielded shorter tests.

Corrected equations. The pattern of correlations observed for the tests using the corrected regression equations paralleled that observed for the classical equations. That is, the range of correlations was fairly small for both the fall and winter quarter data sets, ranging from .818 to .979 under the .01 termination criterion for the fall quarter Energy and Cell subtests, respectively, and from .770 to .887 under the .05 termination criterion for the fall quarter Energy and Chemistry subtests, respectively.

For the winter quarter equations applied to the winter data, the range of conventional-adaptive score correlations was from .889 (for Reproduction) to .983 (for Chemistry) under the .01 criterion and from .715 (for Ecology) to .962 (for Chemistry) under the .05 criterion. In all cases, the correlations obtained using the classical equations were at least as large as, and usually larger than, those obtained using the corrected regression equations.

Cross-Validation

Under the cross-validation conditions (when fall equations were applied to winter data, and vice versa), there was no systematic tendency for the correlations to be either higher or lower than those obtained in the development groups. For the sets of classical and corrected equations alike, cross-validation yielded higher correlations about half the time and lower correlations the other half. Thus, there appears to be no net decrement or increment in the accuracy of measurement when regression equations that were developed on one group were applied in the inter-subtest branching strategy to data for a different group.

Information

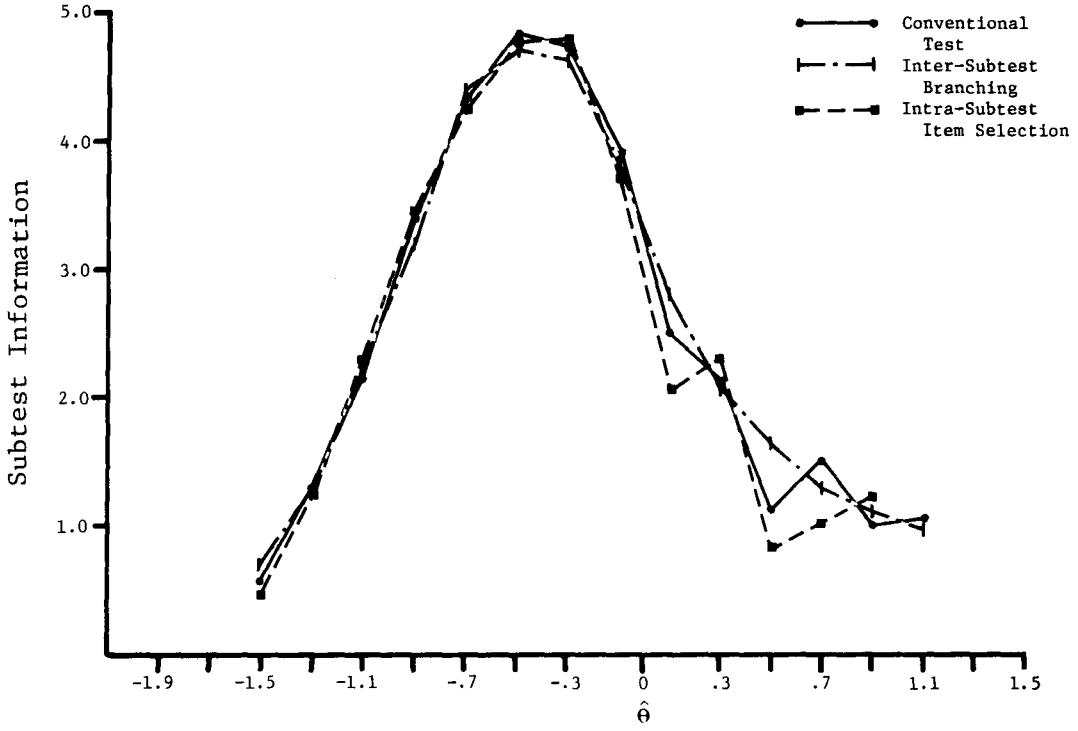
Appendix Tables E through M present the subtest information curves for each subtest under the various testing conditions and across the two academic quarters. It should be noted that since the Chemistry subtest was administered first each quarter (Table E), the initial Bayesian prior $\hat{\theta}$ and variance were 0.0 and 1.0, respectively, for all students over all testing conditions. Thus, because the first subtests administered were identical, there were no differences in the values of the subtest information curves across testing conditions within one termination criterion.

Adaptive Intra-Subtest Item Selection

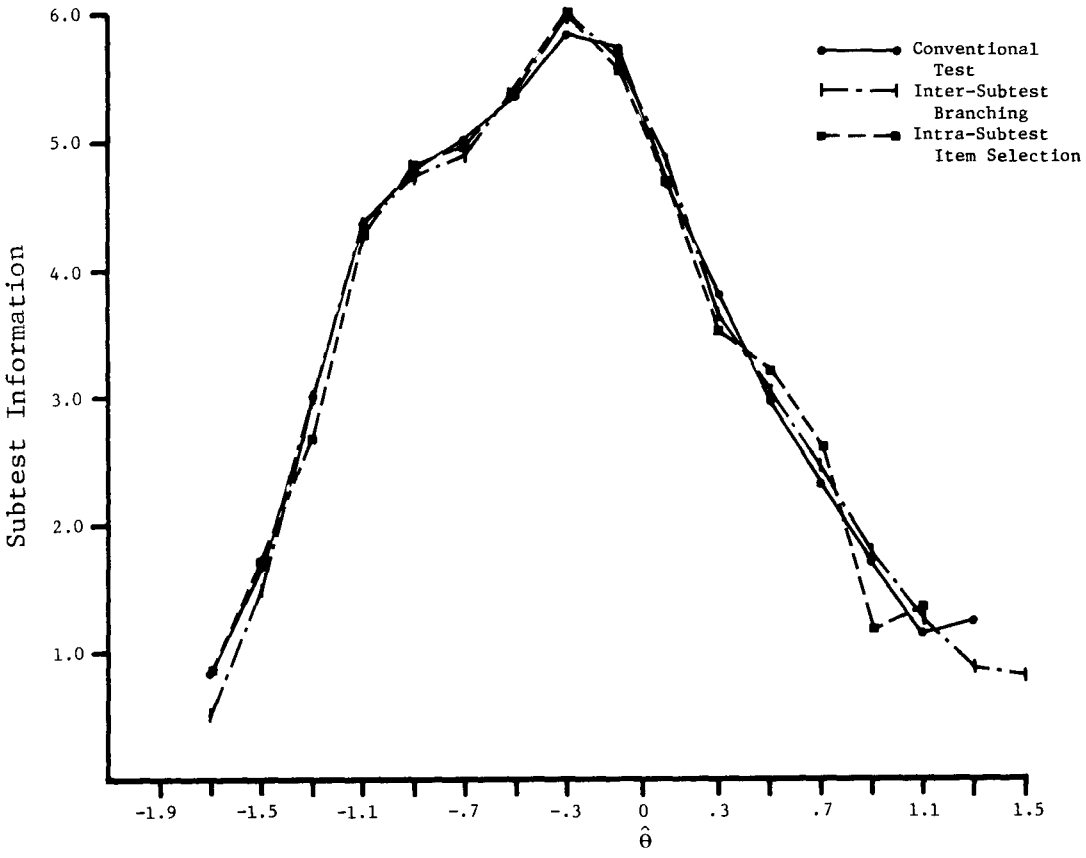
To illustrate the findings with respect to information for the various testing conditions, Figures 1a and 1b present the information curves for the fall quarter Cell and Reproduction subtests (see Tables F and H) obtained when the tests were administered conventionally and with adaptive intra-subtest item selection (termination criterion of .05). The curves are virtually indistinguishable in each case. That is, there was little, if any, loss of information incurred by utilizing an adaptive intra-subtest item selection strategy, even though previous results indicated that the adaptive tests were shorter than the conventional tests.

Figure 1
Subtest Information Curves for the Fall Quarter Cell and Reproduction Subtests Administered Conventionally, with Intra-Subtest Item Selection and Inter-Subtest Branching

(a) Cell



(b) Reproduction



For the Cell subtest (Figure 1a) there was a slightly larger separation between the curves above the point at which the curves were peaked, with the adaptive test slightly lower than the conventional test; this pattern is not evident in Figure 1b. The differences observed in these figures were even smaller when the more stringent termination criterion (.01) was used (see Tables F and H).

Inter-Subtest Branching

Classical equations. Also included in Figures 1a and 1b are the information curves obtained using an inter-subtest branching strategy with the classical fall equations and a termination criterion of .05. There is, again, minimal separation among the curves, particularly for the Reproduction subtest. As before, the curves begin to differ for the Cell subtest in the upper tail, with the inter-subtest branching strategy resulting in higher information values than the other two strategies.

Corrected equations. For both the fall and winter data the information curves obtained using the corrected equations were nearly always lower than the curves obtained with the classical equations. While this difference was small, it was consistent across all five subtests for each quarter (see Tables F through M).

Cross-Validation

When the classical regression equations were used on the fall data, subtest information was slightly, though systematically, higher under cross-validation than for the development groups. That is, applying winter quarter equations to fall quarter data yielded higher levels of information, on the average, than did applying the fall quarter equations to the fall data. This effect was consistent across all five subtests for the fall data. For the winter data, the results were mixed.

When the corrected regression equations were used in cross-validation, the results were mixed for both sets of data. For about half of the subtests, there was a small increase in information, and for the rest of the subtests there was a small decrease in information; thus, there was no net change in information on cross-validating with the corrected equations. In all cases, differences between mean information levels across the various testing conditions were slight.

DISCUSSION

This paper has endeavored to replicate previously reported findings (Brown & Weiss, 1977) that a combination of adaptive intra-subtest item selection and inter-subtest branching strategies could significantly reduce the length of an achievement test battery, with a corresponding minimal loss in psychometric test information. The present study applied this adaptive testing strategy to the responses from a conventionally administered classroom exam and separated out the effects of adaptive intra-subtest item selection and inter-subtest branching on test length and test information. In addition, this paper investigated the effects of using an adaptive testing strategy developed from one set of data on a different data set using a double-cross-validation design.

Adaptive Intra-Subtest Item Selection

The adaptive intra-subtest item selection strategy used in this study was identical to that utilized by Brown and Weiss (1977); that is, items were selected on the basis of the amount of psychometric information available at the current level of $\hat{\theta}$. Although the θ estimates would most appropriately be obtained using a maximum likelihood scoring strategy, this strategy utilized a Bayesian scoring approach. Maximum likelihood scoring requires the availability of at least one correct and one incorrect response before a $\hat{\theta}$ can be generated, and the Bayesian routine has no such requirement. With the possibility of a very small number of items being administered in any one subtest, and the necessity of scoring responses after each item, a maximum likelihood method would be nonoptimal for this testing strategy.

Kingsbury and Weiss (1979) illustrated the extent to which these two scoring methods, when applied to the same set of data, yield scores that are numerically discrepant. The issue of the appropriate choice of scoring strategy pervades implementations of ICC test theory in general and hence is not confined to this particular implementation of an adaptive testing strategy. Nevertheless, it is not known to what extent the results reported here would have changed had the scoring routine been different.

As Table 8 indicates, most of the reduction in test length was due to the variable termination criterion of the intra-subtest item selection strategy. Although test length decreased, the conventional-adaptive test score correlations remained high (often close to 1.00; see Table 10), and there was virtually no loss in the amount of psychometric information available for each subtest. It is clear from these data that subtest length can be reduced from 16% to 32%, with minimal loss in measurement accuracy and precision, simply by omitting those items which add little information to the measurement process.

Inter-Subtest Branching

Utilization of prior information in the estimation of achievement levels further decreased test length by less than 5%, and most often by 1% or less. Although this additional effect was small, it appeared to be fairly consistent across types of regression equations and sets of data; that is, in nearly all cases the addition of the inter-subtest branching strategy resulted in some increased reduction in test length.

Brown and Weiss (1977) reported an average decrease in the length of their test battery of approximately 50%. The largest decrease in the present study was approximately 32%, and that was obtained with a termination criterion (.05) less stringent than the one used in the former study. Part of this discrepancy may lie in the number of items available in each subtest and in the total test. In the earlier study, each subtest was between 12 and 24 items long, and the entire battery contained 201 items. The biology tests used in the present study, however, were much shorter, with a total of only 49 items during fall quarter and 37 items during winter quarter; the lengths of the subtests were correspondingly small. It seems reasonable that the longer subtests in the Brown and Weiss study contained much redundant information and that this would naturally lead to larger reductions in test length.

It would be interesting to compare between studies the extent to which inter-subtest branching reduced test length over and above that obtained by

intra-subtest item selection alone. Unfortunately, Brown and Weiss (1977) did not present that information. More research is needed to determine how representative the present figure of 5% is across different data sets.

When Brown and Weiss computed the conventional-adaptive test score correlations, they found that most of them were above .90, with only 1 of their 12 correlations dropping below that value. There was a greater range for these correlation coefficients in the present study, although here, too, most of them were greater than .90. The lengths of the subtests varied across the two studies, so direct comparison of the correlation coefficients is difficult. The correlations obtained in the previous study may have been larger than in the present one, but the adapted subtests were typically longer as well. This is very likely due to the part-whole correlations which would necessarily increase with the size of the smaller (adapted) part.

Both of these studies concluded that there was minimal loss in the amount of psychometric information observed in each subtest. Brown and Weiss utilized termination criterion of .01 and .001; it is interesting to note that the same conclusion was reached in the present study, which utilized termination criteria that were much less stringent (.05 and .01).

Corrected Regression Equations

The use of Lawley and Maxwell's (1973) correction for error in the independent variables in multiple regression increased the value of the multiple correlation coefficient and the regression coefficients (see Tables 3 and 4). The important issue here, however, was whether this correction affected test length, and accuracy and precision of measurement. On the average, use of the corrected equations decreased test length slightly more than did use of the classical equations. It was impossible to detect any large difference in this data set, however, because there was such a small additional reduction in test length attributable to any kind of inter-subtest branching.

The average correlations between the adaptive and conventional achievement estimates were lower when the corrected equations were used than when the classical equations were used. Although this is puzzling in light of the data in Tables 3 and 4, it becomes less so considering the fact that the corrected equations typically resulted in shorter test lengths. At least part of the discrepancies among the correlation coefficients can be attributed to the discrepancies in test lengths. It is not clear, however, just how much is artifactual and how much is due to a genuine difference in the way the levels of achievement were estimated.

Additionally, mean information values obtained using the corrected regression equations were typically lower than those obtained with the classical equations. At least part of this difference may be attributable to the shorter test lengths that accompanied the corrected equations, although, again, the extent to which this is true is not known.

Cross-Validation

In this study the regression equations for the inter-subtest branching strategies were developed from data from two different academic quarters. These equations were then applied to the data from the other quarter in a

double-cross-validation design to investigate the extent to which the equations, and hence the inter-subtest branching strategies, were sample-specific. This was done for both the classical and corrected sets of equations.

In terms of test length, the cross-validation groups typically were administered shorter tests than were each of the development groups. This was true in nearly all cases under the .01 termination criterion; results were mixed for the .05 criterion.

The accuracy of measurement, as indexed by the correlation between conventional and adaptive test scores, was not systematically affected by the cross-validation procedure employed here. That is, cross-validating yielded higher correlations about half the time and lower correlations the other half, regardless of whether the classical or corrected equations were used. The precision of measurement (i.e., subtest information) increased slightly under cross-validation over that observed for the development groups, at least for the winter quarter and some of the fall sets of classical equations; results were mixed for the corrected equations.

The increases in accuracy and precision of measurement under cross-validation, though slight, are contrary to expectations, since cross-validating yielded shorter mean test lengths as well. Therefore, the increase in measurement accuracy and precision cannot be accounted for by test length changes.

CONCLUSIONS

The real-data simulation reported here replicated and extended the findings reported by Brown and Weiss (1977). That is, the results from this study show that test length could be reduced by 20%-30% using Brown and Weiss's adaptive testing strategy for achievement testing batteries. Reduced time in testing means more time available to be spent in other activities, such as additional instruction.

The level of reduction in test length depended directly on the size of the termination criterion employed. The termination criteria used here were minimum item information of .05 and .01; Brown and Weiss used a value of .01 in their study. Clearly, the choices for termination were arbitrary, and the results might have been different, depending on the value chosen. More research is needed to determine optimal termination criteria.

The design of this study permitted the separation of the effects due to the intra-subtest item selection procedure from those due to inter-subtest branching. Results from this study show that most of the reduction in test length could be attributed to the adaptive intra-subtest item selection method and variable termination criterion. When this strategy was coupled with inter-subtest branching, an additional reduction in test length of only up to 5% was observed. More research is needed to determine the specific characteristics of the item pool which would contribute to greater reductions in test length when the inter-subtest branching strategies are used.

Achievement level estimates obtained adaptively correlated quite highly with those obtained from a conventional administration of the subtests. It is only when the subtests were very short (less than three items) that low correlations were observed.

As was observed in the Brown and Weiss (1977) study, there was a minimal loss in the amount of psychometric information available in the subtests due to adaptive testing. This was evident in the close correspondence between the information curves for the adaptive and conventional tests.

Perhaps the most important finding from this research was that the regression equations obtained from one set of data could be used to adapt the testing for a different group of students and that the observed test characteristics for this cross-validated group closely paralleled the results obtained from the development group. This result directly reflects what would actually happen in a live-testing implementation of this adaptive testing strategy; that is, the regression equations used for inter-subtest branching would be obtained from one group of students and applied in the testing of a different group of students. This study has shown that such a procedure can be utilized while still maintaining the quality of test characteristics observed for the original group on which the regression equations were developed. Of course, more research is needed to determine the generality of these findings in other situations.

Although this study has replicated and extended some of the findings reported by Brown and Weiss (1977), it was limited by the fact that it, too, was a real-data simulation study. The next step in research on this adaptive testing strategy should be the implementation of this adaptive testing strategy in a live-testing situation, thus enabling researchers to evaluate the validity of the findings from these simulation studies. In addition, more research is needed to determine the generality of these findings across other test batteries and other testing situations.

REFERENCES

Bejar, I. I., & Weiss, D. J. A construct validation of adaptive achievement testing (Research Report 78-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1978.

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A047495)

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828)

Betz, N. E., & Weiss, D. J. Simulation studies of two-stage ability testing (Research Report 74-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1974. (NTIS No. A001230)

Betz, N. E., & Weiss, D. J. Empirical and simulation studies of flexilevel ability testing (Research Report 75-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, July 1975. (NTIS No. A013185)

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

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A046062)

Kingsbury, G. G., & Weiss, D. J. Effect of point-in-time in instruction on the measurement of achievement (Research Report 79-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, August 1979.

Larkin, K. C., & Weiss, D. J. An empirical investigation of computer-administered pyramidal ability testing (Research Report 74-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, July 1974. (NTIS No. AD 78553)

Larkin, K. C., & Weiss, D. J. An empirical comparison of two-stage and pyramidal adaptive ability testing (Research Report 75-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1975. (NTIS No. AD A006733)

Lawley, D. N., & Maxwell, A. E. Regression and factor analysis. Biometrika, 1973, 60, 331-338.

Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 117-138.

Maxwell, A. E. Limitations on the use of the multiple linear regression model. British Journal of Mathematical and Statistical Psychology, 1975, 28, 51-62.

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (NTIS No. AD A022964)

Neter, J., & Wasserman, W. Applied linear statistical models: Regression, analysis of variance, and experimental designs. Homewood, IL: Richard D. Irwin, 1974.

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Urry, V. W. A five-year quest: Is computer-assisted testing feasible? In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Personnel Research and Development Center, PS-75-6). Washington, DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-00940-9)

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

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1975. (NTIS No. AD A020961)

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376)

APPENDIX: SUPPLEMENTARY TABLES

Table A
Normal Ogive Item Discrimination (a), Difficulty (b), and Lower Asymptote (c) Parameter Estimates for the Fall Quarter Final Exam, by Subtest

Subtest and Item	a	b	c
Chemistry			
1	1.76	.87	.37
2	1.60	-.68	.27
3	1.39	-1.41	.49
4	1.55	.33	.32
5	.77	-.66	.15
6	1.54	-.56	.30
7	-	-	-
8	1.98	-.78	.28
9	2.36	-.60	.23
10	.92	-.93	.30
11	1.66	-1.57	.36
12	-	-	-
13	1.67	.63	.39
Cell			
1	1.48	.63	.43
2	2.53	3.01	.59
3	1.84	1.68	.49
4	1.79	-.28	.32
5	2.08	-.87	.34
6	1.82	-.70	.40
7	2.26	-.48	.54
8	1.17	.12	.51
9	1.58	-1.02	.41
Energy			
1	2.77	.06	.29
2	1.99	-.83	.59
3	2.01	1.41	.43
4	1.68	-.19	.59
5	1.74	1.10	.38
6	2.73	.45	.22
7	2.04	.36	.40
8	2.93	-1.58	.50
9	2.54	-1.26	.34
Reproduction			
1	1.18	0.00	.46
2	1.69	-.76	.40
3	1.47	.54	.49
4	.73	-.24	.34
5	1.40	2.03	.57
6	2.28	-1.36	.61
7	1.08	-.53	.21
8	2.41	-1.05	.25
9	1.79	-.07	.30
10	2.53	-.33	.24
11	1.52	.38	.53
Ecology			
1	1.58	-1.35	.38
2	1.45	-1.19	.47
3	2.36	-1.64	.55
4	1.66	-.33	.36
5	-	-	-
6	1.91	-.14	.41
7	1.42	-.15	.48

Note. Missing entries indicate that the item was rejected in the first phase of item parameter estimation.

Table B
Normal Ogive Item Discrimination (a), Difficulty (b), and Lower Asymptote (c) Parameter Estimates for the Winter Quarter Final Exam, by Subtest

Subtest and Item	a	b	c
Chemistry			
1	1.76	.87	.37
2	-	-	-
3	2.21	-.82	.16
4	1.60	-.68	.27
5	1.26	.66	.37
6	1.55	.33	.32
7	-	-	-
8	1.54	-.56	.30
9	2.36	-.60	.23
10	1.85	-1.50	.29
Cell			
1	1.48	.63	.43
2	1.45	-.20	.30
3	1.84	1.68	.49
4	2.08	-.87	.34
5	1.48	-1.06	.32
6	1.82	-.70	.40
Energy			
1	-	-	-
2	2.20	1.49	.42
3	2.28	-.05	.49
4	2.85	.92	.33
5	2.07	-.49	.68
6	2.73	.45	.22
7	2.09	-.69	.50
8	1.35	-.17	.48
Reproduction			
1	1.14	-.94	.33
2	-	-	-
3	1.47	.54	.49
4	1.40	2.03	.57
5	-	-	-
6	1.30	-.76	.30
7	2.05	-1.01	.53
8	1.85	1.52	.53
9	1.52	.38	.53
Ecology			
1	1.22	-.46	.38
2	-	-	-
3	1.93	1.92	.79
4	2.28	-1.22	.37

Note. Missing entries indicate that the item was rejected in the first phase of item parameter estimation.

Table C

Regression Coefficients and Multiple Correlation Coefficients (R) for the Intermediate Classical Regression Equations from the Fall and Winter Quarter Final Exams

Quarter and Criterion Subtest	Regression Coefficients for Scores on Previously Administered Subtests				Regression Constant	R
	Chemistry	Cell	Energy	Reproduction		
Fall						
Two Independent Variables						
Energy	.328	.272			-.009	.464*
Reproduction	.286	.228			.203	.434
Ecology	.286	.163			-.392	.415
Three Independent Variables						
Reproduction	.240	.190	.140		.204	.455*
Ecology	.251	.134	.107		-.291	.429
Four Independent Variables						
Ecology	.221	.110	.089	.128	-.029	.446*
Winter						
Two Independent Variables						
Cell	.256		.305		-.144	.525*
Reproduction	.294		.243		.115	.421
Ecology	.140		.085		.120	.244
Three Independent Variables						
Reproduction	.258	.129	.203		.134	.432*
Ecology	.129	.040	.073		.125	.248
Four Independent Variables						
Ecology	.102	.026	.052	.103	.112	.278

Note. An asterisk (*) indicates that the criterion subtest in that particular row was designated as the next subtest to be administered.

Table D

Factor Loadings and Communality Estimates For Maximum Likelihood Factor Analyses of Fall and Winter Quarter Final Exams

Fall Quarter

Two Independent Variables: Criterion Subtest = Energy

$$\Lambda^* \begin{bmatrix} \text{Energy} \\ \text{Chemistry} \\ \text{Cell} \end{bmatrix} = \begin{bmatrix} .594 \\ .693 \\ .624 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .352 \\ .481 \\ .389 \end{bmatrix}$$

Three Independent Variables: Criterion Subtest = Reproduction

$$\Lambda^* \begin{bmatrix} \text{Reproduction} \\ \text{Chemistry} \\ \text{Cell} \\ \text{Energy} \end{bmatrix} = \begin{bmatrix} .552 \\ .698 \\ .623 \\ .590 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .304 \\ .487 \\ .388 \\ .348 \end{bmatrix}$$

Four Independent Variables: Criterion Subtest = Ecology

$$\Lambda^* \begin{bmatrix} \text{Ecology} \\ \text{Chemistry} \\ \text{Cell} \\ \text{Energy} \\ \text{Reproduction} \end{bmatrix} = \begin{bmatrix} .523 \\ .712 \\ .611 \\ .581 \\ .555 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .274 \\ .506 \\ .374 \\ .338 \\ .309 \end{bmatrix}$$

Winter Quarter

Two Independent Variables: Criterion Subtest = Cell

$$\Lambda^* \begin{bmatrix} \text{Cell} \\ \text{Chemistry} \\ \text{Energy} \end{bmatrix} = \begin{bmatrix} .644 \\ .701 \\ .707 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .415 \\ .491 \\ .501 \end{bmatrix}$$

Three Independent Variables: Criterion Subtest = Reproduction

$$\Lambda^* \begin{bmatrix} \text{Reproduction} \\ \text{Chemistry} \\ \text{Energy} \\ \text{Cell} \end{bmatrix} = \begin{bmatrix} .504 \\ .717 \\ .700 \\ .634 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .254 \\ .542 \\ .490 \\ .402 \end{bmatrix}$$

Four Independent Variables: Criterion Subtest = Ecology

$$\Lambda^* \begin{bmatrix} \text{Ecology} \\ \text{Chemistry} \\ \text{Energy} \\ \text{Cell} \\ \text{Reproduction} \end{bmatrix} = \begin{bmatrix} .303 \\ .722 \\ .694 \\ .628 \\ .514 \end{bmatrix} \qquad h^2 = \begin{bmatrix} .092 \\ .522 \\ .481 \\ .394 \\ .264 \end{bmatrix}$$

Table E
Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Chemistry Subtest of the Fall and Winter Quarter Final Exams for the Conventional Test and the Adaptive Test Using Only Intra-Subtest Item Selection with Two Termination Criteria

$\hat{\theta}$ Range		Fall						Winter					
		Conventional		Adaptive Test: Termination Criterion				Conventional		Adaptive Test: Termination Criterion			
				.01		.05				.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	1	1.25	5	1.11	4	.97	-	-	-	-	-	-
-1.799	-1.600	7	1.56	5	1.66	7	1.56	8	1.21	19	1.31	19	1.30
-1.599	-1.400	19	2.34	19	2.22	18	2.23	22	1.91	22	2.08	22	2.08
-1.399	-1.200	25	2.78	33	2.77	33	2.77	33	2.69	31	2.80	31	2.80
-1.199	-1.000	68	3.78	56	3.77	55	3.77	49	4.05	41	4.03	36	3.96
-0.999	-0.800	64	5.28	55	5.22	57	5.18	77	5.69	68	5.64	65	5.54
-0.799	-0.600	86	6.82	79	6.69	67	6.68	61	6.83	60	6.75	64	6.48
-0.599	-0.400	85	6.92	58	6.98	55	6.97	92	6.52	53	6.48	58	6.39
-0.399	-0.200	79	5.97	84	5.93	64	5.98	67	5.40	96	5.28	73	5.27
-0.199	0.000	40	4.53	56	4.55	58	4.63	57	4.05	69	4.14	78	4.35
0.001	0.200	43	3.50	52	3.46	37	3.32	45	2.87	46	2.85	46	2.85
0.201	0.400	42	3.06	32	3.05	36	3.00	43	2.42	44	2.46	59	2.39
0.401	0.600	41	2.90	84	3.09	95	3.05	104	2.41	103	2.42	91	2.44
0.601	0.800	61	3.23	19	3.16	20	3.15	7	1.00	-	-	7	1.00
0.801	1.000	4	1.27	5	1.28	11	1.39	21	1.44	34	1.38	31	1.55
1.001	1.200	47	1.77	64	1.85	170	2.11	114	2.02	114	2.08	120	2.11
1.201	1.400	88	2.05	94	2.16	13	2.20	-	-	-	-	-	-
1.401	1.600	-	-	-	-	-	-	-	-	-	-	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	4.27	800	4.07	800	3.90	800	3.92	800	3.78	800	3.72

Table F

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Cell Subtest of the Fall Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching															
								Classical Equations								Corrected Equations							
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	.09	6	.10
-1.799	-1.600	-	-	-	-	-	-	-	-	-	-	7	.26	7	.24	-	-	-	-	14	.23	12	.23
-1.599	-1.400	5	.59	7	.46	7	.46	3	.66	4	.69	16	.45	16	.49	3	.66	4	.69	18	.55	17	.51
-1.399	-1.200	24	1.28	28	1.34	30	1.27	19	1.22	24	1.25	23	1.30	27	1.23	19	1.22	24	1.25	27	1.14	45	1.10
-1.199	-1.000	42	2.19	34	2.27	31	2.25	31	2.12	27	2.23	34	2.17	31	2.25	31	2.12	27	2.23	34	2.09	18	2.15
-0.999	-0.800	52	3.37	58	3.55	57	3.45	41	3.23	40	3.20	41	3.33	41	3.21	41	3.23	40	3.20	42	3.54	48	3.39
-0.799	-0.600	61	4.30	59	4.26	48	4.24	58	4.31	53	4.37	70	4.40	72	4.38	58	4.31	53	4.37	73	4.34	67	4.44
-0.599	-0.400	119	4.82	103	4.78	79	4.76	100	4.74	81	4.74	100	4.74	85	4.86	100	4.74	81	4.74	72	4.77	71	4.83
-0.399	-0.200	61	4.72	68	4.82	51	4.79	81	4.68	65	4.63	70	4.62	54	4.56	81	4.68	65	4.63	62	4.60	43	4.45
-0.199	0.000	23	3.92	30	3.71	30	3.71	46	3.90	30	3.83	45	3.68	40	3.67	46	3.90	30	3.83	65	3.75	53	3.78
0.001	0.200	25	2.52	12	2.07	12	2.07	53	2.74	57	2.79	75	2.88	74	2.88	53	2.74	57	2.79	65	2.93	63	2.93
0.201	0.400	131	2.17	144	2.32	144	2.32	82	2.20	77	2.16	78	2.15	76	2.14	82	2.20	77	2.16	70	2.20	65	2.16
0.401	0.600	20	1.11	16	1.07	1	.83	76	1.66	65	1.67	87	1.70	76	1.69	76	1.66	65	1.68	87	1.70	74	1.71
0.601	0.800	86	1.50	90	1.49	73	1.04	72	1.41	119	1.30	56	1.41	92	1.33	72	1.41	119	1.30	44	1.43	69	1.34
0.801	1.000	34	1.00	33	1.01	237	1.22	51	1.20	77	1.10	46	1.19	64	1.13	51	1.20	77	1.11	31	1.14	54	1.12
1.001	1.200	117	1.09	118	1.11	-	-	57	1.02	81	.99	52	1.03	45	1.01	57	1.02	81	.99	39	1.01	35	.95
1.201	1.400	-	-	-	-	-	-	30	1.02	-	-	-	-	-	-	30	1.02	-	-	29	.95	60	.94
1.401	1.600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	1.02	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	2.72	800	2.73	800	2.46	800	2.76	800	2.51	800	2.80	800	2.64	800	2.76	800	2.51	800	2.67	800	2.49

Table G
Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Energy Subtest of the Fall Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching													
								Classical Equations								Corrected Equations					
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	2	1.16	4	1.10	-	-	1	1.17	-	-	5	.61	5	.60	8	.70	5	.57	5	.61	5	.60
-1.799	-1.600	16	2.17	6	1.34	10	1.23	16	2.02	15	2.03	8	2.28	8	2.25	26	2.13	29	2.05	8	2.31	8	2.30
-1.599	-1.400	42	3.44	37	3.37	27	3.30	37	3.33	31	3.21	29	3.56	27	3.54	23	3.31	18	3.22	29	3.56	27	3.54
-1.399	-1.200	28	4.01	25	4.07	21	4.06	14	3.92	12	3.68	25	3.99	19	3.98	25	3.90	21	3.80	25	3.99	19	3.98
-1.199	-1.000	33	3.50	23	3.37	7	3.14	31	3.55	30	3.50	25	3.43	19	3.33	28	3.61	26	3.53	25	3.43	19	3.33
-0.999	-0.800	57	2.67	40	2.87	47	3.03	44	2.67	37	2.68	42	2.67	39	2.72	50	2.62	46	2.64	42	2.67	39	2.72
-0.799	-0.600	74	2.07	55	2.05	54	2.05	58	1.99	44	2.02	43	1.98	42	1.98	53	2.04	44	2.04	43	1.98	42	1.98
-0.599	-0.400	90	1.73	106	1.73	126	1.69	72	1.74	104	1.70	79	1.75	93	1.69	74	1.74	88	1.70	79	1.75	93	1.69
-0.399	-0.200	66	2.11	56	2.29	52	2.31	73	2.11	60	2.06	83	2.12	77	2.04	60	2.15	55	2.14	83	2.12	77	2.04
-0.199	0.000	65	3.51	50	3.26	51	3.24	68	3.48	59	3.43	73	3.45	73	3.64	68	3.49	62	3.52	73	3.45	73	3.64
0.001	0.200	79	5.41	96	5.30	137	5.53	80	5.46	88	5.15	79	5.27	78	5.23	63	5.29	69	5.05	79	5.27	78	5.23
0.201	0.400	43	6.59	61	6.28	11	3.91	43	6.36	31	6.39	54	6.44	44	6.09	47	6.11	33	6.50	54	6.44	44	6.09
0.401	0.600	13	3.83	25	5.21	28	4.90	41	5.94	51	5.71	43	6.09	49	5.87	49	6.06	58	5.90	43	6.09	49	5.87
0.601	0.800	24	4.62	38	4.52	43	4.52	34	4.22	29	4.30	40	4.26	38	4.45	30	4.65	30	4.60	40	4.26	38	4.45
0.801	1.000	41	3.62	19	3.36	20	3.89	27	3.31	34	3.32	26	3.27	27	3.21	28	3.11	28	3.18	26	3.27	27	3.21
1.001	1.200	18	1.70	24	1.68	24	1.68	24	2.09	20	2.03	47	2.20	46	2.17	32	2.30	29	2.31	47	2.20	46	2.17
1.201	1.400	41	2.07	42	1.86	54	2.01	73	2.26	91	2.19	56	2.40	110	2.42	41	2.37	59	2.18	56	2.41	110	2.42
1.401	1.600	68	2.51	93	2.47	88	2.46	58	2.30	58	2.38	43	2.51	6	2.50	49	2.07	48	2.01	43	2.51	6	2.50
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	39	1.72	45	1.82	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	3.18	800	3.28	800	3.17	794	3.28	794	3.18	800	3.38	800	3.28	793	3.22	793	3.15	800	3.38	800	3.28

Table H

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Reproduction Subtest of the Fall Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching													
								Classical Equations								Corrected Equations					
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	1	.13	1	.37	3	.16	3	.32	1	.21
-1.799	-1.600	10	.85	4	.82	5	.85	1	.98	2	.53	6	.86	6	.72	9	.75	7	.73	7	.61	9	.63
-1.599	-1.400	11	1.68	14	1.72	11	1.70	16	1.55	14	1.51	16	1.54	15	1.55	16	1.64	14	1.59	22	1.49	19	1.52
-1.399	-1.200	30	3.01	23	2.77	25	2.68	17	2.94	20	2.97	24	3.11	23	3.06	23	3.02	23	3.00	23	3.06	26	3.17
-1.199	-1.000	44	4.33	48	4.30	51	4.25	38	4.32	39	4.29	40	4.30	40	4.20	35	4.24	38	4.18	33	4.18	31	4.12
-0.999	-0.800	43	4.78	43	4.82	29	4.79	43	4.70	33	4.74	33	4.75	28	4.73	28	4.86	24	4.86	32	4.85	24	4.84
-0.799	-0.600	33	5.01	25	5.00	18	4.97	26	5.06	15	4.88	30	5.08	18	5.06	30	5.03	17	5.07	27	4.99	20	5.01
-0.599	-0.400	26	5.34	20	5.27	22	5.37	27	5.47	23	5.42	22	5.45	21	5.46	31	5.53	27	5.53	31	5.53	29	5.51
-0.399	-0.200	44	5.84	66	5.98	62	6.02	59	6.01	51	5.98	66	5.94	60	5.91	57	5.93	46	5.85	66	5.90	56	5.91
-0.199	0.000	71	5.75	71	5.66	67	5.63	54	5.64	60	5.71	60	5.68	58	5.72	59	5.82	69	5.80	57	5.73	64	5.76
0.001	0.200	90	4.66	92	4.66	103	4.67	81	4.82	84	4.80	81	4.77	88	4.75	75	4.75	77	4.78	71	4.82	72	4.81
0.201	0.400	85	3.81	64	3.51	66	3.53	95	3.69	104	3.68	89	3.67	98	3.67	99	3.70	111	3.70	100	3.71	108	3.67
0.401	0.600	130	3.01	126	3.20	127	3.22	109	3.03	104	3.02	105	3.05	102	3.03	85	3.02	72	2.98	79	3.10	75	3.05
0.601	0.800	11	2.36	20	2.44	19	2.60	57	2.48	56	2.48	53	2.49	46	2.51	82	2.38	78	2.39	81	2.40	74	2.41
0.801	1.000	52	1.70	57	1.67	23	1.19	43	1.82	55	1.78	45	1.79	59	1.78	43	1.85	47	1.88	36	1.83	51	1.79
1.001	1.200	55	1.15	127	1.22	172	1.36	58	1.38	78	1.26	53	1.40	79	1.28	32	1.36	55	1.30	40	1.39	53	1.28
1.201	1.400	65	1.24	-	-	-	-	53	.97	36	.87	55	.94	32	.87	47	1.00	27	.98	40	1.02	23	.94
1.401	1.600	-	-	-	-	-	-	23	.78	26	.82	22	.80	26	.82	22	.66	40	.66	30	.68	40	.68
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	26	.63	25	.60	22	.63	23	.63
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	3.59	800	3.61	800	3.55	800	3.52	800	3.42	800	3.53	800	3.44	800	3.46	800	3.36	800	3.47	800	3.38

Table I

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Ecology Subtest of the Fall Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item-Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching													
								Classical Equations								Corrected Equations					
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	11	1.05	3	.97	3	.97	4	.81	4	.81	1	1.13	1	1.12	5	1.05	3	.97	-	-	1	1.27
-1.799	-1.600	25	1.86	10	1.72	10	1.72	8	1.80	7	1.79	8	1.75	6	1.68	12	1.79	11	1.79	4	1.53	3	1.61
-1.599	-1.400	24	2.20	34	2.33	23	2.30	25	2.30	25	2.31	19	2.35	20	2.34	18	2.27	21	2.27	15	2.34	16	2.35
-1.399	-1.200	16	2.40	15	2.37	13	2.35	23	2.36	22	2.33	21	2.42	17	2.34	27	2.37	25	2.37	24	2.41	16	2.41
-1.199	-1.000	42	2.15	30	2.15	30	2.14	29	2.17	28	2.14	28	2.18	28	2.15	33	2.21	33	2.15	22	2.17	18	2.12
-0.999	-0.800	24	1.94	40	1.97	35	1.98	37	2.00	33	2.00	40	1.97	37	1.97	38	2.01	37	1.99	38	1.99	36	1.99
-0.799	-0.600	77	2.01	83	2.03	81	2.03	76	1.95	70	1.95	80	1.99	75	2.00	72	1.92	59	1.93	62	1.99	54	2.00
-0.599	-0.400	41	2.03	35	2.01	20	1.86	55	2.27	55	2.78	52	2.18	55	2.22	56	2.31	58	2.30	61	2.20	64	2.18
-0.399	-0.200	148	2.66	148	2.63	158	2.55	104	2.69	100	2.65	150	2.68	153	2.64	90	2.68	85	2.64	94	2.66	92	2.63
-0.199	0.000	32	2.98	32	2.96	33	2.92	63	2.82	67	2.80	31	2.83	33	2.82	68	2.77	71	2.78	100	2.85	105	2.82
0.001	0.200	-	-	-	-	-	-	29	2.22	27	2.25	20	1.76	16	1.77	45	2.47	45	2.47	13	2.43	11	2.51
0.201	0.400	1	.55	2	.71	10	.78	64	1.76	67	1.78	114	1.75	135	1.71	61	1.83	68	1.83	46	1.65	49	1.63
0.401	0.600	103	.90	107	.92	104	.98	106	1.30	108	1.24	208	1.36	199	1.32	78	1.31	70	1.22	122	1.29	136	1.25
0.601	0.800	254	.97	261	.95	280	1.03	109	.85	108	.83	28	1.07	25	1.06	64	.84	69	.81	140	.88	133	.83
0.801	1.000	-	-	-	-	-	-	68	.55	79	.55	-	-	-	-	73	.56	71	.53	59	.59	66	.59
1.001	1.200	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46	.31	65	.30	-	-	-	-
1.201	1.400	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	.22	6	.25	-	-	-	-
1.401	1.600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		798	1.70	800	1.69	800	1.67	800	1.80	800	1.77	800	1.95	800	1.93	797	1.78	797	1.73	800	1.80	800	1.76

Table J

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Cell Subtest of the Winter Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching													
								Classical Equations								Corrected Equations					
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	.18	-	-
-1.799	-1.600	-	-	-	-	-	-	-	-	-	-	4	.33	2	.38	-	-	-	-	13	.34	16	.32
-1.599	-1.400	15	.68	21	.67	21	.67	4	.52	4	.57	24	.69	25	.70	4	.52	4	.57	36	.65	26	.64
-1.399	-1.200	38	1.39	34	1.31	34	1.31	39	1.29	38	1.30	43	1.19	43	1.17	39	1.29	38	1.30	35	1.25	44	1.17
-1.199	-1.000	51	2.12	64	2.26	64	2.26	39	2.06	38	2.06	48	2.16	48	2.15	39	2.06	38	2.06	52	2.20	49	2.17
-0.999	-0.800	83	3.17	68	3.16	67	3.17	71	3.08	71	3.06	66	3.11	65	3.10	71	3.08	71	3.06	61	3.05	61	3.01
-0.799	-0.600	68	3.52	65	3.48	56	3.48	65	3.45	59	3.45	78	3.43	66	3.44	65	3.45	59	3.45	74	3.51	65	3.58
-0.599	-0.400	66	3.30	69	3.25	53	3.47	80	3.26	69	3.21	92	3.40	97	3.32	80	3.26	69	3.21	86	3.38	86	3.30
-0.399	-0.200	81	3.00	81	2.99	98	2.88	96	2.88	99	2.90	84	2.81	82	2.89	96	2.88	99	2.90	79	2.81	81	2.85
-0.199	0.000	102	2.40	94	2.38	94	2.38	75	2.19	73	2.19	78	2.19	80	2.18	75	2.19	73	2.19	83	2.14	76	2.20
0.001	0.200	59	1.67	104	1.70	104	1.70	91	1.78	88	1.79	77	1.70	78	1.71	91	1.78	88	1.79	70	1.74	68	1.71
0.201	0.400	42	1.53	1	.24	1	.24	48	1.38	53	1.39	60	1.41	58	1.42	48	1.38	53	1.39	62	1.42	65	1.45
0.401	0.600	71	1.21	80	1.14	15	.69	56	1.14	45	1.11	57	1.20	62	1.19	56	1.14	45	1.11	43	1.22	42	1.19
0.601	0.800	32	.82	27	.85	193	1.08	50	1.11	97	1.06	41	1.14	59	1.08	50	1.11	97	1.06	35	1.10	61	1.06
0.801	1.000	92	1.02	92	1.03	-	-	34	1.01	26	.84	20	1.02	18	.91	34	1.01	26	.84	29	1.00	19	.92
1.001	1.200	-	-	-	-	-	-	52	.90	40	.89	28	.92	17	.91	52	.90	40	.89	11	.86	10	.81
1.201	1.400	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	29	.90	31	.87
1.401	1.600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	2.18	800	2.15	800	2.12	800	2.13	800	2.08	800	2.18	800	2.15	800	2.13	800	2.08	800	2.11	800	2.07

Table K

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Energy Subtest of the Winter Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching															
								Classical Equations								Corrected Equations							
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-1.799	-1.600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	.01	3	.01	-	-	-	-
-1.599	-1.400	-	-	-	-	-	-	10	.06	4	.05	2	.06	-	-	39	.04	30	.03	2	.06	-	-
-1.399	-1.200	20	.18	24	.16	3	.21	33	.14	31	.13	17	.15	16	.18	31	.14	39	.13	17	.15	16	.18
-1.199	-1.000	55	.47	51	.42	72	.33	64	.42	56	.35	43	.42	42	.36	70	.43	63	.37	43	.42	42	.36
-0.999	-0.800	87	.94	77	.86	77	.86	72	.91	87	.86	81	.90	83	.90	63	.92	74	.90	81	.90	83	.91
-0.799	-0.600	88	1.53	85	1.45	80	1.44	73	1.57	74	1.58	67	1.56	67	1.56	78	1.57	71	1.57	67	1.56	67	1.56
-0.599	-0.400	111	2.21	91	2.09	56	2.00	91	2.13	93	2.18	88	2.17	91	2.18	72	2.16	81	2.18	88	2.17	91	2.18
-0.399	-0.200	55	2.28	65	2.40	147	2.44	68	2.51	70	2.49	84	2.53	87	2.50	62	2.51	62	2.48	84	2.53	87	2.50
-0.199	0.000	93	2.82	133	2.91	77	2.82	85	2.68	59	2.68	98	2.71	70	2.76	76	2.61	54	2.61	98	2.71	70	2.76
0.001	0.200	54	3.11	34	2.58	34	2.58	48	2.91	64	3.04	58	3.00	81	3.04	44	2.99	58	2.90	58	3.00	81	3.04
0.201	0.400	14	3.00	80	4.12	79	4.15	68	4.15	54	3.95	61	3.99	48	3.80	50	4.16	42	4.03	61	3.99	48	3.80
0.401	0.600	70	4.50	18	1.82	19	1.78	46	4.23	50	4.22	59	4.21	59	4.31	51	4.42	53	4.56	59	4.21	59	4.31
0.601	0.800	19	2.40	28	4.62	36	4.37	24	3.67	27	3.55	26	4.18	30	3.65	35	4.17	39	3.91	26	4.18	31	3.70
0.801	1.000	35	3.74	11	2.12	10	2.16	25	3.78	25	3.80	29	3.45	28	3.96	22	3.92	19	3.89	29	3.45	27	3.92
1.001	1.200	43	2.86	46	2.79	53	3.13	31	2.95	32	3.38	33	3.62	44	3.39	24	3.44	26	3.74	33	3.62	44	3.39
1.201	1.400	56	3.02	18	3.02	18	3.02	33	2.79	45	2.74	38	2.80	54	2.96	32	2.56	34	2.54	38	2.80	54	2.96
1.401	1.600	-	-	39	2.77	39	2.77	29	2.49	29	2.53	16	2.77	-	-	21	2.02	35	2.06	16	2.77	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24	1.94	17	1.99	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	2.34	800	2.30	800	2.31	800	2.32	800	2.33	800	2.51	800	2.52	800	2.22	800	2.22	800	2.51	800	2.52

Table L

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Reproduction Subtest of the Winter Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching															
								Classical Equations								Corrected Equations							
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	.22	2	.22	2	.20	1	.20
-1.799	-1.600	16	.43	16	.43	16	.43	2	.38	2	.38	11	.37	10	.37	10	.35	8	.34	14	.32	15	.32
-1.599	-1.400	4	.44	4	.44	4	.44	22	.61	22	.61	20	.61	21	.60	22	.55	25	.55	25	.57	24	.57
-1.399	-1.200	54	1.13	57	1.14	57	1.14	23	1.09	21	1.11	31	1.08	28	1.09	27	1.11	23	1.15	25	1.10	23	1.12
-1.199	-1.000	30	1.59	31	1.72	29	1.74	31	1.60	33	1.59	25	1.66	28	1.61	27	1.62	30	1.61	36	1.63	37	1.60
-0.999	-0.800	72	2.04	21	1.82	15	1.65	45	1.99	27	1.94	65	2.03	47	2.02	58	2.03	50	2.02	63	2.04	45	1.98
-0.799	-0.600	82	2.06	116	2.08	131	2.05	75	2.10	77	2.05	67	2.04	78	2.02	75	2.06	66	2.00	67	2.05	76	2.04
-0.599	-0.400	42	1.76	33	1.77	12	1.71	66	1.83	69	1.81	60	1.85	51	1.81	56	1.81	63	1.81	51	1.83	56	1.78
-0.399	-0.200	62	1.59	70	1.60	70	1.60	61	1.51	56	1.51	64	1.48	64	1.50	72	1.52	63	1.51	73	1.51	62	1.51
-0.199	0.000	81	1.25	25	1.08	25	1.08	84	1.31	81	1.30	87	1.32	83	1.32	74	1.31	73	1.31	78	1.32	75	1.32
0.001	0.200	92	1.30	147	1.31	147	1.31	82	1.28	75	1.28	70	1.28	69	1.28	75	1.27	67	1.27	67	1.28	60	1.28
0.201	0.400	5	.94	9	.90	2	.54	62	1.32	60	1.31	55	1.32	57	1.30	63	1.32	69	1.30	62	1.31	72	1.30
0.401	0.600	56	1.26	38	1.32	57	.89	34	1.29	51	1.28	37	1.27	61	1.30	35	1.32	40	1.31	33	1.29	50	1.32
0.601	0.800	86	1.25	119	1.22	235	1.33	67	1.30	72	1.26	74	1.29	72	1.25	56	1.30	53	1.27	58	1.33	52	1.27
0.801	1.000	118	1.21	114	1.23	-	-	50	1.15	45	1.12	45	1.17	28	1.09	40	1.19	39	1.13	43	1.14	31	1.10
1.001	1.200	-	-	-	-	-	-	47	1.09	93	1.01	47	1.09	85	1.02	35	1.05	71	.99	31	1.08	73	1.00
1.201	1.400	-	-	-	-	-	-	38	1.11	9	1.15	30	1.10	12	1.16	32	1.12	17	1.12	32	1.11	16	1.09
1.401	1.600	-	-	-	-	-	-	11	1.29	7	1.22	12	1.28	6	1.21	23	1.15	28	1.13	24	1.17	19	1.12
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	1.28	13	1.28	16	1.30	13	1.30
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	1.44	800	1.43	800	1.42	800	1.43	800	1.38	800	1.42	800	1.39	800	1.42	800	1.38	800	1.41	800	1.38

Table M

Mean Information Values ($\bar{I}$) at Estimated Achievement Level ($\hat{\theta}$) Intervals for the Ecology Subtest of the Winter Quarter Final Exam Under all Testing Conditions

$\hat{\theta}$ Range		Conventional Test		Adaptive Intra-Subtest Item Selection: Termination				Adaptive Intra-Subtest Item Selection with Inter-Subtest Branching															
								Classical Equations								Corrected Equations							
								Fall: Termination				Winter: Termination				Fall: Termination				Winter: Termination			
				.01		.05		.01		.05		.01		.05		.01		.05		.01		.05	
Lo	Hi	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$	N	$\bar{I}$
-2.000	-1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	.18	-	-	-	-	-	-
-1.799	-1.600	-	-	-	-	-	-	6	.47	5	.53	-	-	-	-	9	.32	14	.35	-	-	-	-
-1.599	-1.400	21	.86	21	.83	21	.83	19	1.00	18	.94	11	1.07	10	1.07	27	1.00	23	1.01	4	1.12	4	1.13
-1.399	-1.200	79	1.52	79	1.48	-	-	51	1.67	52	1.66	52	1.72	53	1.72	34	1.61	35	1.60	24	1.61	24	1.62
-1.199	-1.000	-	-	-	-	-	-	21	1.98	2	1.91	37	1.96	24	1.94	29	1.98	12	1.97	68	1.97	33	1.95
-0.999	-0.800	-	-	-	-	-	-	11	1.69	11	1.68	-	-	-	-	14	1.71	14	1.71	1	1.92	-	-
-0.799	-0.600	44	1.16	44	1.16	44	1.16	20	1.29	20	1.29	27	1.27	28	1.26	26	1.23	25	1.19	11	1.19	11	1.19
-0.599	-0.400	-	-	-	-	-	-	23	.82	20	.80	16	1.01	15	1.01	51	.91	51	.92	26	.98	25	.99
-0.399	-0.200	-	-	-	-	-	-	49	.64	51	.65	18	.35	16	.31	76	.73	75	.73	14	.49	14	.54
-0.199	0.000	-	-	-	-	-	-	118	.59	116	.59	154	.57	147	.57	113	.60	101	.60	50	.49	51	.48
0.001	0.200	27	.07	26	.05	26	.05	175	.48	212	.46	311	.49	367	.47	114	.48	136	.47	186	.47	242	.45
0.201	0.400	629	.38	630	.37	709	.38	146	.37	134	.35	164	.38	140	.36	100	.37	102	.34	221	.38	210	.35
0.401	0.600	-	-	-	-	-	-	111	.28	105	.27	10	.32	-	-	88	.28	83	.27	152	.28	144	.27
0.601	0.800	-	-	-	-	-	-	50	.20	54	.20	-	-	-	-	62	.19	73	.20	43	.20	42	.21
0.801	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	36	.14	41	.14	-	-	-	-
1.001	1.200	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19	.11	15	.11	-	-	-	-
1.201	1.400	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.401	1.600	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.601	1.800	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.801	2.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total Group		800	.54	800	.52	800	.42	800	.62	800	.57	800	.67	800	.64	800	.63	800	.58	800	.59	800	.51

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

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

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 CDR Robert S. Kennedy
Naval Aerospace Medical and Research Lab
Box 29407
New Orleans, LA 70189

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 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
Naval Health Research Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION & TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs (Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc (OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

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

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 LCOL Gary Ploedorn
Training Effectiveness Analysis Division
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, NM 88002

1 DR. RALPH DUSEK
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. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. 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 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

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

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

Marines

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

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM
Ft. 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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

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

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Warner Birice
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State Collcge
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
NAVPERS/Western Division
27255 Zerwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Michael Levine
Department of Educational Psychology
University of Illinois
Champaign, IL 61820

1 Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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

1 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

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

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

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

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

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

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

1 Dr. Robert Tsutakawa
Dept. of Statistics
University of MIssouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Varicl Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. John Wannous
Department of Management
Michigan University
East Lansing, MI 48824

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979. (NTIS No. AD A068176)
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979. (NTIS No. AD A067928)
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979. (NTIS No. AD A067752)
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978. (NTIS No. AD A064352)
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978. (NTIS No. AD A062797)
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978. (NTIS No. AD A059436)
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977. (NTIS No. AD A047495)
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977. (NTIS No. AD A046062)
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977. (NTIS No. AD A044828)
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977. (NTIS No. AD A041090)
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977. (NTIS No. AD A041084)
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977. (NTIS No. AD A038114)
Final Report: Computerized Ability Testing, 1972-1975. April 1976. (NTIS No. AD A024516)
76-5. Effects of Item Characteristics on Test Fairness. December 1976. (NTIS No. AD A035393)
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976. (NTIS No. AD A027170)
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976. (NTIS No. AD A028147)
76-2. Effects of Time Limits of Test-Taking Behavior. April 1976. (NTIS No. AD A024422)
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (NTIS No. AD A022964)
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975. (NTIS No. AD A020961)
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (NTIS No. AD A018675)
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975. (NTIS No. AD A018758)
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (NTIS No. AD A013185)
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975. (NTIS No. AD A007572)
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (NTIS No. AD A006733).
74-5. Strategies of Adaptive Ability Measurement. December 1974. (NTIS No. AD A004270)
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974. (NTIS No. AD A001230)
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (NTIS No. AD 783553)
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (NTIS No. AD 781894)
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974. (NTIS No. AD 773961)
73-3. The Stratified Adaptive Computerized Ability Test. September 1973. (NTIS No. AD 768376)
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973. (NTIS AD 757788)

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.
Copies of these reports are available, while supplies last, from:
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455

The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models

Tom E. Trabin
and
David J. Weiss

RESEARCH REPORT 79-7
DECEMBER 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Navy Personnel Research and Development Center, and Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 79-7	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* The Person Response Curve: The Fit of Individuals to Item Characteristic Curve Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Tom E. Trabin and David J. Weiss		8. CONTRACT OR GRANT NUMBER*(s)* N00014-76-C-0243
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N PROJ.:RR042-04 T.A.:RR042-04-01 W.U.:NR150-382
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE December 1979 13. NUMBER OF PAGES 36
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 in 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

This research was supported by funds from the Navy Personnel Research and Development Center, and the Office of Naval Research, and monitored by the Office of Naval Research.

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

intra-individual dimensionality	carelessness	achievement testing
three-parameter logistic	guessing	adaptive testing
latent trait test theory	testwiseness	tailored testing
item characteristic curve theory	ability testing	person-fit
	deviant responses	

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

This study investigated a method of determining the fit of individuals to item characteristic curve (ICC) models using the person response curve (PRC). The construction of observed PRCs is based on an individual's proportion correct on test item subsets (strata) that differ systematically in difficulty level. A method is proposed for identifying irregularities in an observed PRC by comparing it with the expected PRC predicted by the three-parameter ICC logistic model for that individual's ability level. Diagnostic

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

potential of the PRC is discussed in terms of the degree and type of deviations of the observed PRC from the expected PRC predicted by the model.

Observed PRCs were constructed for 151 college students using vocabulary test data on 216 items of wide difficulty range. Data on students' test-taking motivation, test-taking anxiety, and perceived test difficulty were also obtained. PRCs for the students were found to be reliable and to have shapes that were primarily a function of ability level. Three-parameter logistic (ICC) model expected PRCs served as good predictors of observed PRCs for over 90% of the group. As anticipated from this general overall fit of the observed data to the ICC model, there were no significant correlations between degree of non-fit and test-taking motivation, test-taking anxiety, or perceived test difficulty. Using split-pool observed PRCs, a few students were identified who deviated significantly from the expected PRC.

The results of this study suggested that three-parameter logistic expected PRCs for given ability levels were good predictors of test response profiles for the students in this sample. Significant non-fit between observed and expected PRCs would suggest the interaction of additional dimensions in the testing situation for a given individual. Recommendations are made for further research on person response curves.

CONTENTS

Acknowledgments

The assistance of Isaac Bejar, Austin T. Church, J. Stephen Prestwood, and James B. Sympson in various aspects of the design of this study and the analysis of these data is gratefully acknowledged.

Technical Editor: Barbara Leslie Camm

The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models

The development of group ability tests more than 50 years ago has enabled the comparison of the total test score of an individual with the scores of a population norm group, thus allowing for more meaningful interpretation of ability estimates than can be done with the use of simple number-correct scores. For example, the statement "On the XYZ aptitude test John scored at the 73rd percentile of college students" gives more information than the statement,"On the ABC ability test Mary correctly answered 64 questions out of 90, whereas Sam correctly answered 33 questions." Both examples have in common the report of a person's test performance on a specific dimension given in terms of an overall test score; but this single summary score, while more parsimonious than a description of a testee's entire response pattern, may not reveal the operation of other factors on test-taking behavior, such as guessing, anxiety, cultural bias, or lack of motivation. Thus, total scores on a test do not indicate whether that test is inappropriate for a certain individual or group of individuals.

The emergence of modern test theory, based on the item characteristic curve (ICC; Hambleton & Cook, 1977; Lord & Novick, 1968), brings with it the promise of better tests conveying more accurate information about testee ability levels. This is partially accomplished by use of ability estimation procedures that take into account the testee's total response pattern in estimating ability levels (Bejar & Weiss, 1979; Kingsbury & Weiss, 1979a). These scoring methods can provide individualized error bands around the testee's ability level estimates, which indicate the precision of those ability estimates (e.g., Kingsbury & Weiss, 1979b). Thus, in addition to providing methods designed to permit more adequate test construction by the use of test information curves (Hambleton & Cook, 1977), ICC theory permits utilizing more of a testee's response pattern in order to provide individualized estimates of precision for ability estimates. In addition, ICC theory also allows for the development of powerful methods of adaptive testing for the solution of many practical measurement problems (e.g., Brown & Weiss, 1977; Kingsbury & Weiss, 1979b; McBride & Weiss, 1976; Vale & Weiss, 1977; Weiss, 1973, 1975).

In contrast to classical test theory, ICC theory makes strong assumptions in order to achieve its objectives. The major operational forms of ICC theory assume 1) local independence, 2) unidimensionality, and 3) a specified shape for the item characteristic curve. Although local independence cannot be directly demonstrated, data supporting the unidimensionality assumption in a variety of settings (e.g., Bejar, Weiss, & Kingsbury, 1977; Church, Pine, & Weiss, 1978; Martin, Pine, & Weiss, 1978; McBride & Weiss, 1974; Reckase, 1978) lend indirect support to the assumption of local independence. Lord (1968) has presented data showing that the assumption of a normal ogive ICC is tenable and, given the minor differences between a logistic ogive and a normal ogive, has indirectly supported the use of the logistic item response function in ICC theory.

There has been very little research, however, to demonstrate that *individuals* behave in accordance with the ICC model, although a growing concern has been exhibited in the testing literature for the development of methods to extract more information from test response data than simply a total score. Use of ICC models with individuals must rest on a demonstration that the test responses of individuals are in accordance with the testing model hypothesized. If this can be demonstrated for most individuals on a number of tests, ICC models can be used with confidence to their full power. On the other hand, if a majority of individuals respond in ways contrary to ICC theory, the utility of the theory for individual measurement can be seriously questioned.

A major advantage of the assumptions of ICC theory for individual measurement is that the question of individuals' fit or non-fit to the model can be investigated on an individual basis. The practical implications of identifying non-fitting persons were realized by Educational Testing Service in their study of methods to identify response patterns of the type of student who "may be so atypical and unlike other students that his [or her] aptitude test score fails to be a completely appropriate measure of his [or her] relative ability" (Levine & Rubin, 1976). Examples of such students are low-ability examinees who copy answers to several difficult items from a much more able neighbor and very high-ability examinees fluent in another language but not yet fluent in English, who misunderstand the wording of several relatively easy questions. Levine and Rubin recommended the development of indices to identify such test item response patterns as a "rich and fertile area for future research."

The appropriateness of a certain test or certain items for specific individuals has also been an important concern for test developers working with the one-parameter logistic ICC (Rasch) model. Wright and his associates (1977; Mead, 1979; Wright & Stone, 1979; Wainer & Wright, in prep.) have proposed identification of such factors as guessing, carelessness, and bias, using the Rasch model. According to Lumsden (1977), a bright but careless student may have the same overall ability score as a careful and consistent average student, but there are differential instructional implications for teaching these two types of students or differential counseling implications if the two students are seeking vocational counseling.

Thus, the question of fit or non-fit of individuals to ICC testing models has important practical and theoretical importance. Fit of individuals must be demonstrated in order to realize the full potential of the model for practical use. At the same time, the development of reliable and valid methods for quantifying and identifying aberrant response patterns would provide a potentially useful source of additional information on test-taking behavior of individuals.

Related Research

The question of fit of individuals to the ICC models can be conceptualized as investigating the variability of a single individual in a single testing situation. Wright (1977), in suggesting that to postulate and to study such a phenomonon would be to "wreak havoc with the logic and practice of measurement," exemplifies an attitude which may, in part, account for the meager literature on the topic. It is more likely, however, that the development of sufficiently refined measurement techniques to handle such a difficult problem has not occurred until very recently. The development of computerized testing together with the development of latent trait test theory was necessary to bring about the possibility of measuring individual variablity with a single test.

Most of the existing research consists of tentative theoretical approaches with closing exhortations for further study. The approaches to this problem differ widely in theoretical orientation and in terminology used. Mosier (1942) first referred to individual variability in mental test theory from a psychophysical orientation; and Levine and Rubin (1976) referred to aberrance indices from the view of signal detection theory. Lumsden (1977) used the Thurstoneian approach of categorical judgment to propose the idea of person reliability. Weiss (1973) used data from adaptive testing to develop consistency scores, and Vale and Weiss (1975) further developed the earlier idea of consistency scores into an empirical study of subject characteristic curves. Wright (1977) used the one-parameter logistic model to propose the idea of item residuals and to refute the notion of what he called person sensitivity in testing. Clearly, the idea is still new, hazily formulated on a theoretical level, with very scarce evidence of any empirical studies.

Mosier's psychophysical approach. The first reference in the testing literature to an individual's variability within a single ability testing situation was in a two-part study by Mosier (1940, 1942). The emphasis in this study was on the fundamental relationships between the field of mental test theory and the methods of measuring psychophysical processes. This comparison included relating the constant method of psychophysical measurement with scoring by the number-correct method in mental testing. Mosier asserted that a composite score is an imperfect representation of an individual's test score and depends on the individual's variability, just as an individual's threshold in psychophysics depends on the ambiguity of the stimulus; as a stimulus is variable with respect to a group of judges, so an individual is variable with respect to a group of items.

Mosier likened the ambiguity (discriminal dispersion) of a stimulus in psychophysics to individual variability in mental test performance. He postulated the distribution of the proportion of correct answers for one individual across items of differing difficulty as the integral of the normal probability curve and the variability of that individual as the standard deviation of the probability function whose integral is the proportion of correct answers as a function of difficulty. Mosier applied the constant process of psychophysics to a set of test data (of unspecified characteristics) and estimated the difficulty of median error for individuals (ability level) and its dispersion. He found odd-even reliability of ability level estimated by this method to be .88. The reliability of the person variability index was .55, a value significantly different from zero. It was perhaps this apparent low reliability estimate which was responsible for a complete lack of research on person variability for the next 30 years.

Weiss's stradaptive "trace line". The idea of person variability within one test was independently developed by Weiss (1973) as a by-product of computerized adaptive testing. In the design of the stratified-adaptive (stradaptive) test, he ordered ability test items by difficulty levels into strata. In examining testee performance on stradaptive tests, Weiss noted that individuals who correctly answered items of the same average difficulty level differed in terms of the proportion of items they answered correctly at different difficulty levels.

To examine differences in individual variability, Weiss proposed the concept of a "trace line" for a testee's item responses, with items divided into strata of increasing difficulty on the x-axis and proportion correct for an individual

on each stratum on the y-axis, duplicating the suggestion of Mosier 30 years earlier. Weiss hypothesized as did Mosier, that proportion correct would decrease as stratum difficulty increased. Also echoing Mosier, he proposed that the steepness of the slope be interpreted as an index of the consistency of an individual's item responses and the capability of the item pool to discriminate an individual's ability level. The point of inflection of the curve, where 50% of the items were answered correctly (for free-response items) was proposed as an indicator of the difficulty of the item pool for an individual or the position of that individual on the trait continuum. To operationalize the concept of person variability, or what Weiss called "consistency," he suggested calculating several indices, including the standard deviation of item difficulties answered correctly and the standard deviation of item difficulties encountered.

Vale and Weiss (1975) empirically studied some aspects of individual "consistency" as part of a larger study of computer-administered adaptive testing. Included in this study was a test of the hypothesis that more consistent individuals--those with smaller errors of measurement in Mosier's (1940, 1942) formulation--would have more stable ability estimates. The five operationalizations of consistency originally proposed by Weiss (1973) were studied as moderators in the prediction of test-retest reliability of ability estimates. The standard deviation of item difficulties encountered significantly moderated the stability of ability estimates in the expected direction as, to a lesser extent, did the standard deviation of item difficulties answered correctly.

In addition, Vale and Weiss (1975) studied the test-retest reliability of the "trace line" plots for individuals and introduced the new term "subject characteristic curve" for these trace lines. They used canonical redundancy analysis (Weiss, 1972) on the proportion-correct-within-strata data (i.e., the subject characteristic curves) in a retest situation. The results indicated a high degree of predictability of subject characteristic curves on one test from the test scores on the other; redundancies indicated from 47% to 67% common variance across the two testing times. These results indicated a good degree of stability in the proportion of correct responses within the strata of the stradaptive test as indexed by the subject characteristic curves.

Lumsden's subject characteristic curve. The subject characteristic curve was again independently proposed by Lumsden (1977, 1978) as a derivation from Thurstone's law of categorical judgment. Lumsden proposed an attribute-based model of test performance in which a person's ability fluctuates in trends (long-term developmental changes), swells (short-term mood swings), and tremors (moment-to-moment shifts). He assumed tremors to be rapid, random, and normally distributed shifts of ability occurring from moment to moment within a single test situation: The discriminal dispersion of item difficulties stays at zero, and it is only person ability that fluctuates. If the momentary location of a person's ability level is higher than the point location of the item's difficulty, the person will answer an item correctly. If ability is lower at any moment than the item difficulty location, the person will answer that item incorrectly. Lumsden then extended the idea to the plot, for a single person, of item responses at different difficulty levels, which he called the "person characteristic curve." He suggested that the person characteristic curve is "perfectly analogous to the item characteristic curve." Lumsden's basic assumptions, however, are different from the ICC theory assumptions underlying item characteristic curves; an ICC assumes that ability level is constant, not fluctuating,

but that the response to a given test item includes a random error component causing observed item responses to fluctuate around true ability level.

Levine and Rubin's aberrancy indices. Other approaches to the study of intra-individual variability within a test have concentrated on the use of intra-individual variability for test validation rather than on individual ability assessment. Levine and Rubin (1976) and Levine (1979) initiated several studies concerned with individuals or groups of individuals for whom a given test might be invalid and/or inappropriate. Among the populations of concern were those who obtain higher scores because of cheating and those who obtain lower scores because of lack of proficiency in English. Levine and Rubin developed several types of "aberrance indices" to determine at greater than chance level, without reference to demographic data, examinees for whom a given test would be inappropriate.

Their basic assumption was that an aberrant examinee's response pattern to items of varying difficulty should have a low marginal probability, since it is unlikely that a high-ability examinee would incorrectly answer an easy item or a low-ability examinee correctly answer a difficult item. Marginal probability was operationally defined as the average of the conditional probabilities of a correct response on each item of difficulty level b for an individual of ability level θ. If n = the number of items, there are 2^n marginal probabilities. These were ranked, with all probabilities below an arbitrary cutoff point considered to represent aberrant response patterns.

Using a monte carlo simulation with 3,000 hypothetical examinees, 200 of whom were aberrant responders, Levine and Rubin (1976) conducted several studies at different cutoff points on the marginal probabilities to determine if aberrant examinees could be identified at a rate significantly greater than chance. Receiver operator curves (ROC) from signal detection theory were used to evaluate the performance of their experimental methods of identifying aberrance. The best method identified 50% of the spuriously low and 80% of the spuriously high examinees, while only mistaking 10% of the normal examinees as aberrant.

When compared to the chance level predictions of only 10% of spuriously high or low examinees identified while mistaking 10% of the normal examinees, this study seemed to have yielded results that merit further study. However, a closer look reveals the impracticality of Levine and Rubin's best method. Even if the aberrance indices identified 80% of the aberrant examinees (160 out of 200) and only misclassified 10% of the non-aberrant examinees (280 out of 2,800), this would still result in eliminating as invalid the test results of 280 non-aberrant examinees. Levine and Rubin seemed to completely ignore this problem in their paper.

Wright's residual analysis. Wright's (1977; Wright & Mead, 1977; Wright & Stone, 1979) concern with intra-individual variability in a single situation focuses on the interaction of a person with specific test items. Wright has developed methods for identifying items which may be invalid for a certain person or group of persons and which can then be excluded from consideration when calculating ability estimates from those items. Wright (1977; Wright & Stone, 1979, pp. 165-180) cited tendencies such as guessing, cheating, "sleeping" (getting bored with a test and answering later items in a more haphazard fashion), "fumbling" (e.g., answering earlier items with difficulty because of confusion with test format), and cultural bias. Wright's method (Wright & Stone, 1979; Mead,

in prep.) utilizes standardized squares of the residuals between an item's difficulty level and a person's ability level after fitting the one-parameter logistic model to the test data. If these residuals indicate a significantly low probability of responding correctly or incorrectly and the person responded in that way, the tester then has reason to suspect that the item or item set may be invalid for that particular person.

This approach is consistent with Wright's use of the one-parameter Rasch model, which recognizes only a difficulty level of items but not a discrimination parameter or a guessing parameter. Following the assumptions of this model, Wright maintained that the probability of success on more difficult items should always be less than on easier items no matter who attempts the items, so the test developer must prevent variation in item discrimination sufficient to produce item characteristic curves that cross. Also, following this logic, a higher ability person should have a better chance for success no matter what the difficulty of the item attempted, so the test developer must prevent variation in person sensitivity; the result is that person characteristic curves must not cross each other. Wright claimed that the practical problem of variation in item discrimination and person sensitivity can be treated through supervision rather than estimation, using residuals and deleting inappropriate items from a person's responses without interfering with estimates of a person's ability. Wright's method seems to oversimplify response data by ignoring the effects of item discrimination and guessing, as well as precluding the possibility of more subtle diagnoses of added dimensions acting as moderator variables in the testing situation.

Summary and Objectives

The limited literature on person variability within a test thus seems to have three major trends: 1) the direct analysis of person variability as originally suggested by Mosier, later called the testee's trace line by Weiss and subject characteristic curve by Vale and Weiss and the person characteristic curve by Lumsden (1977); 2) designation of highly variable persons as aberrant by Levine and Rubin; and 3) the elimination of aberrant person-item interactions by Wright. Careful analysis of these three approaches indicates that the first approach (that of the person characteristic curve) is the most general of the three, subsuming the other two as special cases: If the entire pattern of a testee's responses is studied as a function of difficulty level of the items, the identification of aberrant response patterns or person-item restrictions follows directly. In addition, postulating a person characteristic curve in conjunction with ICC theory provides a means of testing for single individuals, whether their response patterns fit the theory regardless of the number of parameters assumed.

The purpose of this study was to further explore the Mosier-Weiss-Lumsden idea of the person characteristic curve, to determine its utility as a means of describing testee response variability, and to study the fit of individuals to the ICC model. To emphasize that the curve is derived from the responses of an individual to a set of test items, it was renamed the "person response curve." The focus of this research is on the investigation of the reliability and other psychometric characteristics of the person response curve.

The Person Response Curve

Observed Person Response Curves

Figure 1 is a plot of person response curves (PRCs) for each of three hypothetical testees. To obtain these plots, a number of items of different difficulty levels are administered to a testee. For each difficulty level, the proportion of items answered correctly is plotted as a function of difficulty level. The resulting PRC is representative of one person's performance on one test.

Figure 1
Observed Person Response Curves for Three Hypothetical Persons with the Same Ability Level ($\hat{\theta}$=0.0)

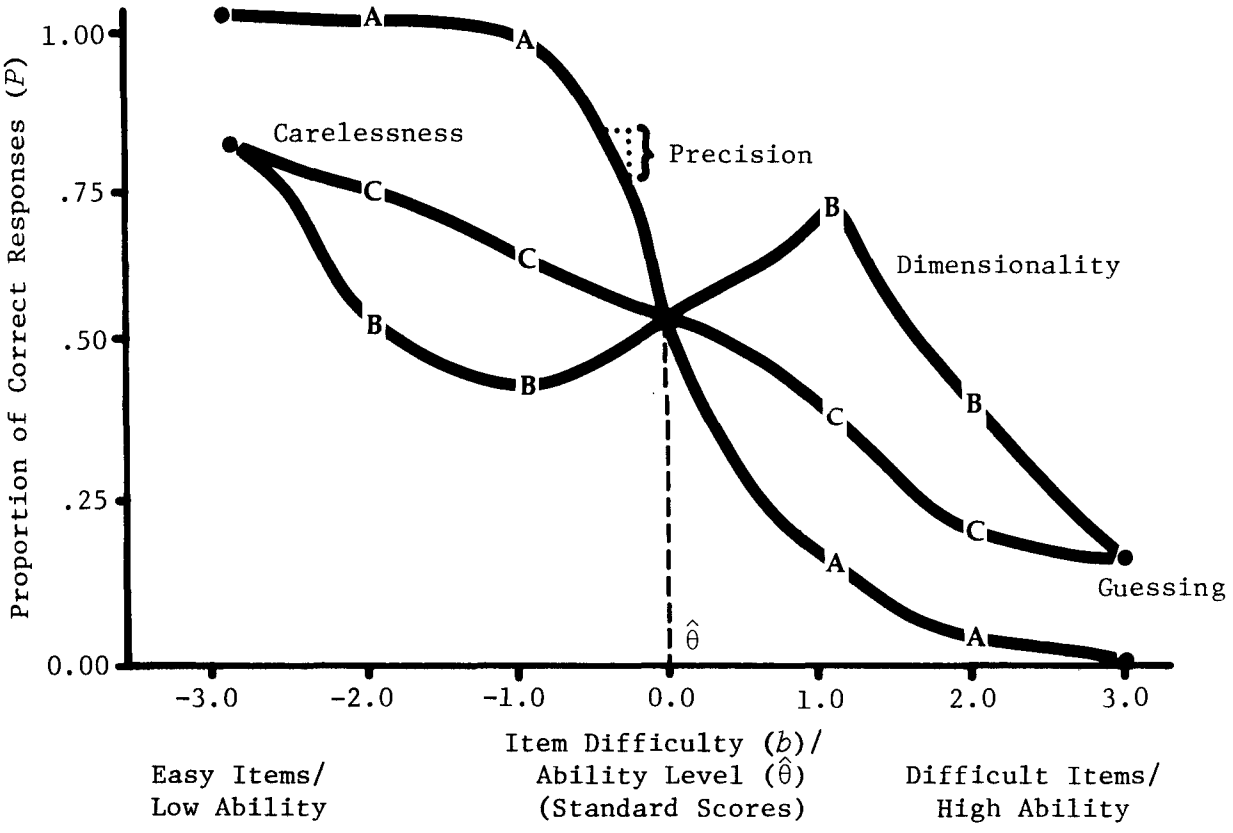


Figure 1 shows the PRC plots of three different persons--A, B, and C--who have all obtained the same score on the test by answering 50% of the total test questions correctly. Thus, all the curves cross at the point on the vertical axis of .50, and their average proportion correct across all item difficulty levels is .50. The center point of the curve can then be projected downward to the horizontal axis to obtain an ability level estimate ($\hat{\theta}$) of 0.0, which in standard score terms is at the mean of a population. Yet, Figure 1 illustrates

that although these three persons all achieved the same total score on this test, they obtained that score in substantially different ways.

As shown in Figure 1, the three testees--A, B, and C--differ in a number of variables. Note that the curve for Person A has a substantially steeper slope around its center point than does that for Persons B and C. This shows that with this particular item pool, Person A was measured more precisely than either Person B or C, or (in Mosier's, 1942, terms) that the error of measurement for Person A was smaller. Thus, in addition to ability level scores, information on individual precision of measurement is derivable from the PRC.

The third type of information derivable from the study of PRCs is a person's guessing behavior. This is shown in Figure 1 as the lower right-hand portion of the curve for each testee. Note that Persons B and C correctly answered very difficult items at a nonzero level. It may, therefore, be hypothesized that they were guessing. However, Person A answered none of the difficult items correctly. It may be hypothesized that this testee, unlike the other two, was not guessing.

A fourth type of information possibly derivable from the PRC is a carelessness index, shown in the upper left-hand corner of Figure 1. Persons B and C answered only about 80% of a set of very easy items correctly, even though their ability levels were considerably higher. On the other hand, Person B answered the same items all correctly, as would be expected for a person with a relatively high ability level. Thus, it could be hypothesized that Persons B and C were more careless than Person A.

Finally, the fifth kind of potential information derivable from a study of PRCs is shown for Person B and is a deviation from a unidimensional response pattern, as suggested by Mosier (1940, p. 364). That is, the test performance of Person B shows that he/she was answering correctly beyond the chance level some difficult items which were beyond his/her ability level. Since such test response behavior is inconsistent with a unidimensional hypothesis, there may be, for this individual, some dimension accounting for test performance other than the one being measured by the test for other persons.

Thus, the PRC provides the potential for considerable additional information from an individual's test response record. All that is required to obtain an observed PRC is 1) to administer to an individual a number of items of varying difficulty levels, 2) to determine the proportion of items answered correctly at each difficulty level, and 3) to plot those proportions as a function of item difficulty level.

Expected Person Response Curves

Although the observed PRCs are useful in describing a person's test behavior, by themselves they provide no means of determining whether observed fluctuations in the curve represent important characteristics of the individual or merely chance deviations. ICC theory, however, permits the derivation of *expected* PRCs, which can then be used to evaluate whether aspects of the observed PRCs are real or chance fluctuations. In addition, these observed PRCs permit testing the fit of individual persons to the ICC model for a given set of test item responses.

Expected PRCs are derivable from either the one-, two-, or three-parameter ICC models. Derivation of the expected PRC requires an ability estimate, $\hat{\theta}$,

and the item parameters for all the items administered. Generally, the ICC item parameters of the items administered will have been estimated in advance by a method such as Lord's LOGIST (Wood, Wingersky,& Lord, 1978) or one of Urry's (e.g., Schmidt & Urry, 1976) estimation procedures; the difficulty (b) parameters will have been used to order the items by difficulty level to obtain the observed PRC. Estimates of ability level ($\hat{\theta}$) may be obtained using programs described by Bejar and Weiss (1979).

In the case of the three-parameter logistic ICC model, the expected probability of a correct response for any given test item (P_g) is given as a function of $\hat{\theta}$, a, b, and c by the three-parameter logistic equation:

$$P_g(\hat{\theta}) = c_g + (1 - c_g)\frac{Da_g(\hat{\theta} - b_g)}{1 + e^{Da_g(\hat{\theta} - b_g)}}, \qquad [1]$$

where

$\hat{\theta}$ is the person's estimated ability score:
g is an item;
a_g is the ICC item discrimination parameter;
b_g is the ICC item difficulty parameter;
c_g is the ICC item lower asymptote ("guessing") parameter' and
D is equal to 1.7.

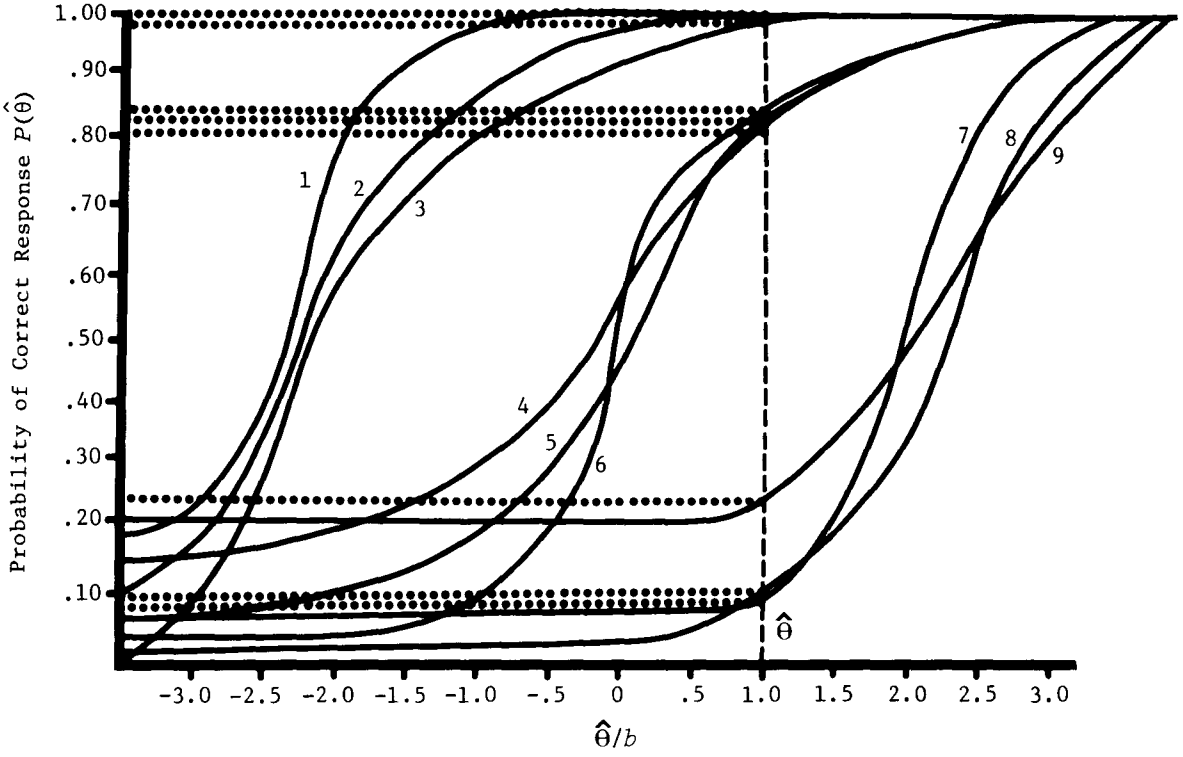
If a two-parameter ICC model is used, the terms in Equation 1 with c are deleted; if the one-parameter (Rasch) model is used, the a values are set to 1.0.

Using the estimated probability of a correct response for each item resulting from Equation 1, an expected PRC can be plotted. This is illustrated in Figure 2. Figure 2a illustrates three-parameter ICCs for nine test items, grouped at three levels of difficulty. Difficulties of Items 1, 2, and 3 are relatively low, between -2.0 and -2.5; Items 4, 5, and 6 are clustered around a difficulty of b=0.0; and Items 7, 8, and 9 are the most difficult set, with $b\simeq$+2.0. The dashed vertical line in Figure 2a represents a person with a $\hat{\theta}$=1.0.

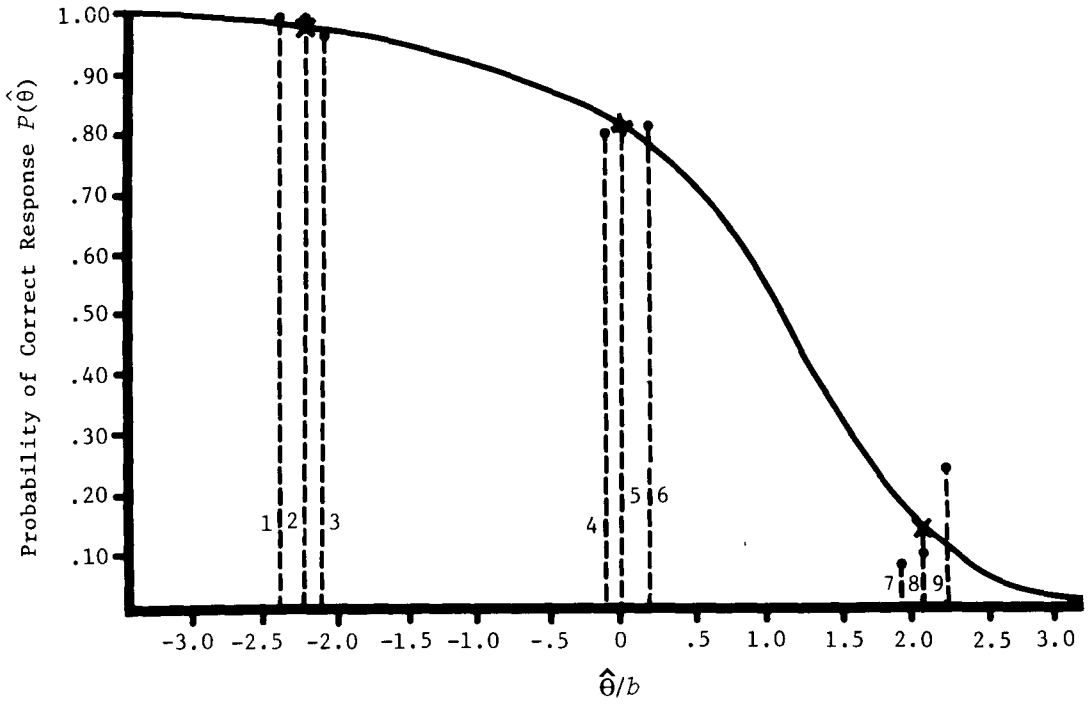
The estimated probability of a correct response to each item, resulting from Equation 1, is shown in Figure 2a by the dotted horizontal line extending from the ICC to the vertical axis at $\hat{\theta}$=1.0. Thus, for Items 1 and 2, the probability of a correct response is essentially 1.0; and for Item 3, about .98. For Items 4, 5, and 6 the probabilities are .80, .82, and .85, respectively; and for Items 7, 8, and 9, P = .08, .10, and .22. These nine probabilities are plotted in Figure 2b and constitute an expected PRC for a person with $\hat{\theta}$=1.0, with the probability for each item plotted at its difficulty level. It will be noted that for Item Groups 4, 5, 6 and 7, 8, 9 in Figure 2a, the expected proportions correct are not monotonically decreasing as might be expected from theoretical considerations. This is due to the differing discriminations of the items (as illustrated in Figure 2a). Thus, to construct an estimated PRC, it might be desirable to plot a smoothed curve around the values plotted in Figure 2b.

Figure 2
Estimating the Expected Person Response Curve (PRC) for a Person with $\hat{\theta}$=1.0 Using Nine Test Items

(a) Three-Parameter Item Characteristic Curves Grouped at Three Levels of Difficulty



(b) Expected Person Response Curve (PRC) for a Person with $\hat{\theta}$=1.0



One way of smoothing expected PRCs is to average the probabilities of a correct response to items close in difficulty level. Since the observed PRC utilizes the proportion of correct responses to a set of items of similar difficulty, averaging of the probabilities of correct responses in the expected PRC will facilitate the direct comparison of observed and expected PRCs. Lord has referred to

$$\zeta = \sum_{g=1}^{k} P_g(\theta) \qquad [2]$$

as the expected true score on a set of test items, where k is the number of items for which the expected probability of a correct response has been computed from Equation 1 and ζ is the expected number of correct responses in k items. An estimate of the proportion of correct responses on a subset of items is

$$\hat{p}_s = \zeta/k = \sum_{i=1}^{k} P_g(\theta)/k \qquad [3]$$

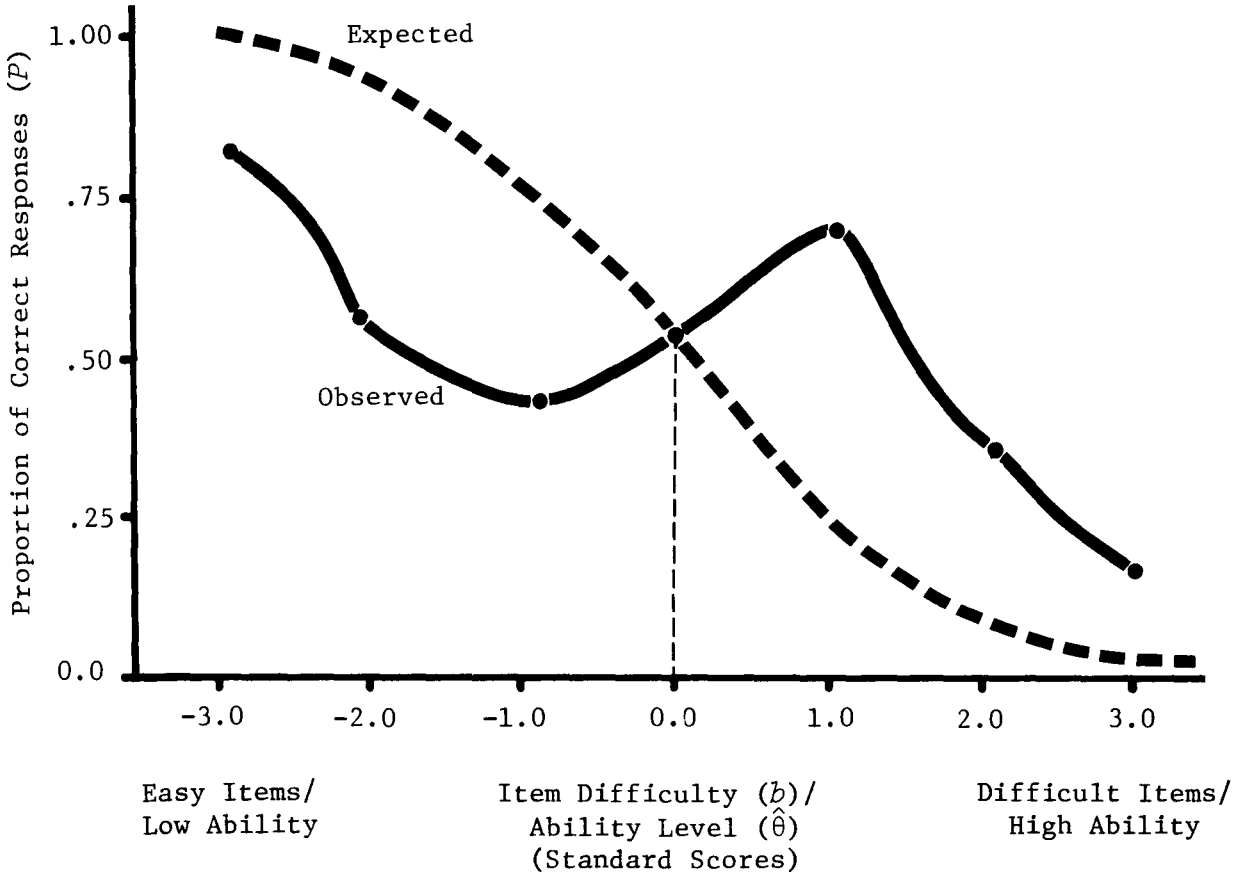
or the average proportion correct on the k-item subset. Values of $\hat{p}_s$, the expected proportion correct on the three subsets of items in Figure 2a, are shown by X's in Figure 2b. Connecting these values with a curve gives the expected PRC based on the three-parameter logistic ICC model, which for any individual is directly comparable to his/her observed PRC.

The expected PRC is therefore simply a function of $\hat{\theta}$ and the item parameters. Thus, for a given $\hat{\theta}$ and a given set of items, the expected values of the PRC will be constant. The observed PRC, on the other hand, results from the interaction of an individual with the items. If an individual answers the set of test items strictly in accordance with the ICC model, the observed PRC should conform to the expected PRC. If an individual's test item responses are determined by factors other than a single unidimensional trait, deviations of the observed PRC from the expected PRC will appear.

Observed versus Expected PRCs

Figure 3 shows hypothetical observed and expected PRCs for an individual with $\hat{\theta}$=0.0. The observed PRC (solid line) is plotted from data on test items grouped at seven points on the item difficulty continuum: b=±3, ±2, ±1, and 0. The expected PRC data points (dashed line) were derived from Equations 1 and 3 for the test items administered, using the same item difficulty groupings. To determine whether a person's carelessness, guessing, dimensionality, or precision are significantly different from those predicted by the model, an expected PRC may be determined for any person on any set of test items with estimated ICC parameters, and the observed PRC may be compared to it. If the observed PRC differs from the expected model-based prediction in any respect, the observed PRC describes a significant aspect of the person's testing behavior. Once quantified, these person-fit variables might then be usable in prediction situations to increase the accuracy of predictions made from test scores. This could be done by including additional information on guessing, carelessness, precision, and dimensionality and on other aspects of a person's test performance as reflected in the relationship of observed and expected PRCs.

Figure 3
Observed and Expected Person Response Curves for a Person with $\hat{\theta}$=0.0



Method

The following data analyses constitute a first examination of observed PRCs and their relationships with expected PRCs for a group of individuals on a test designed to permit study of the characteristics of PRCs. The major analyses were directed at establishing the reliability of observed PRCs and the fit of observed and expected PRCs. Some correlates of person-fit indices derived from the PRC were also investigated.

Subjects

Subjects were 151 undergraduate students in the introductory psychology course at the University of Minnesota. These students volunteered for the study in return for bonus points that would count toward their final grade. Students were given a posttest debriefing, which consisted of a brief explanation of the purpose of the study. No test results were given, due to the lengthy procedures for keypunching and scoring the data.

Test Instrument

The test consisted of 216 five-option multiple-choice vocabulary items. The items were chosen from a preexisting item pool of over 500 items with ICC difficulty and discrimination parameters that had been developed on a similar population of undergraduates in the introductory psychology course in previous years (McBride & Weiss, 1974). The 216 items were selected for high discriminating power and for spread of difficulty (c parameters were set at .20 for all items).

The test was given as a paper-and-pencil test without time limits. Items were randomly ordered for administration so that easy and difficult items were spread throughout the test. In addition, to control for any effects of item order, the pages of test questions were ordered in six different ways so that only one-sixth of the students took the test in the same page order.

Observed PRCs

Stratifying the test. In order to transform student response data into observed PRCs, test items were divided into strata containing an equal number of items, with each stratum representing a different level of difficulty. This was done by reordering the items by difficulty level (b parameter), then dividing them into nine separate groups (or strata) of 24 items each. In this way, Stratum 1 contained the 24 easiest items and Stratum 9 contained the 24 most difficult items.

Items were then ordered within each stratum by discrimination (a) level, with the most discriminating item the first item in the stratum and the least discriminating item the 24th item in the stratum. To investigate the parallel forms reliability of observed PRCs, each stratum was then split into two parallel substrata of items with similar difficulty and discrimination parameters. This provided 18 substrata of 12 items each. Item difficulty and discrimination parameters for all items by stratum and substratum are in Appendix Table A.

Items were scored as either correct ("1") or incorrect ("0"), with omitted items scored as incorrect. The correct-incorrect response vectors were then reordered by item difficulty level for each student. The proportion of correct responses was then computed on each of the nine strata and on each of the 18 substrata for each student, providing information for observed PRCs based on all 216 items (i.e., nine 24-item subtests of differing difficulty levels) and split-half parallel observed PRCs, each based on nine 12-item subtests.

To examine the characteristics of the items constituting the strata, internal consistency reliability of each of the nine strata was determined using Cronbach's alpha. Parallel forms reliability of the nine pairs of parallel substrata was determined by the product-moment correlation coefficient between proportion-correct scores on each of the nine pairs of substrata.

Estimated PRCs

Using Program LINDSCO (Bejar & Weiss, 1979), Owen's Bayesian ability estimation procedure was used to compute ability estimates ($\hat{\theta}$) for each student

based on his/her responses to all 216 items in the test. This $\hat{\theta}$ was then used in Equations 1 and 3, in conjunction with the item parameters for the 24 items in each stratum, to obtain the expected proportion-correct score in each of the nine strata ($\hat{p}_s$). The $\hat{p}_s$ values then constituted the expected PRC for each student, assuming the three-parameter ICC model. This process was repeated for each of the parallel substrata, yielding expected PRCs for each student from each of the two 108-item parallel pools.

Correlates of Observed PRCs

In addition to the vocabulary items, 11 five-alternative Likert-type questions were used to assess psychological variables hypothesized to be related to PRC data. These questions were taken from psychological reactions scales developed by Betz and Weiss (1976), with some slight modifications. Four items were used in a Perceived Test Difficulty scale, four in a Test-Taking Anxiety scale, and three items in a Test-Taking Motivation scale.

The psychological reactions scale items (shown in Appendix Table B) were scored "1" through "5," with the first response alternative for each item scored as "1" and each succeeding alternative scored a point higher. Item scores were weighted positively or negatively (see Table B), according to how they were keyed on the psychological reactions scale. The total number of item score points ranged from +8 to -8 on the Perceived Test Difficulty and the Test-Taking Anxiety scales, and from +9 to -3 on the Test-Taking Motivation scale.

Reliability of Observed PRCs

Within- and between-persons D^2 indices. To determine the split-half parallel forms reliability of observed PRCs, a D^2 statistic was computed for each student, comparing his/her observed PRC data (proportion correct) on each of the paired substrata; thus, D^2 indexed the similarity of the two split-half PRCs for each student. A D^2 value of zero would indicate that the two split-half PRCs were identical; large values would indicate differences between the two PRCs.

Although the D^2 statistic is a commonly used descriptive statistic in comparing profiles (Cronbach & Gleser, 1953), no sampling distribution is available for it. In order to obtain some data with which to compare the split-half D^2 data, four other sets of between-persons D^2 statistics were computed for comparison purposes with the within-persons reliability D^2. Students were paired randomly into 75 pairs. The first D^2 statistic [D(AA)] was obtained by comparing the observed PRC data for one of the split-half PRCs (arbitrarily designated "A") of each individual student with those of his/her randomly paired student. The second D^2 statistic [D(BB)] was obtained by comparing the same pairs on their observed PRC data from their other (Subset B) substrata. The third and fourth D^2 statistics [D(AB) and D(BA)] were obtained by comparing one student's first split-half PRC with the other student's second split-half observed PRC.

Group means and standard deviations were then computed for the four between-persons D^2 indices [D(AA), D(BB), D(AB), and D(BA)] and the one within-persons D^2. Fisher's t test for differences in means was used to determine if within-persons split-half observed PRCs on parallel forms of the test were more similar to each other than they were to the between-persons D^2 from randomly selected individuals. If observed PRC data were reliable, it would be expected that profiles within persons would have significantly lower mean D^2 values than profiles

between persons, especially considering differences in ability level between randomly paired individuals.

Chi-square tests of independence. A second approach to the study of the reliability of observed PRCs used a chi-square test of independence. For each student, the 2 × 9 contingency table included the number of correct responses on each of the parallel substrata in each of the rows of the 9-column table. Chi-square tests of independence were computed separately for each student. If the paired substrata were parallel, a nonsignificant value of chi-square would be supportive of the reliability of observed PRC data. Although this chi-square test violated the usual assumption of independence because the cell frequencies were based on the same student's responses to all the questions, it may be argued that the students' test item responses are locally independent (i.e., are independent for a given student who has a fixed value of θ) and, therefore, that the test is not inappropriate. Further study of this problem is necessary, however, in future applications of this index.

PRCs and Person-Fit

Observed versus expected PRCs. Expected PRCs were determined for each student using the method described above. To determine if students' responses to these ability test items were consistent with the three-parameter ICC model, a chi-square goodness-of-fit statistic was computed between each student's observed and expected PRC data across the nine strata. If the PRC is an adequate index of model fit, the mean chi-square for the group would be nonsignificant. On an individual level, at an .05 level of significance, chi-square goodness-of-fit values should be statistically significant for 7.55 of the 151 students by chance alone, assuming the null hypothesis of no significant deviations from person-fit. More significant chi-square values would indicate a tendency for lack of fit in these data.

When the overall level of fit in the data is substantially different from the chance expectation, it is still difficult to conclude from the overall goodness-of-fit tests that a specific individual exhibited reliable and meaningful lack of fit to the ICC model, since a certain number of such deviations from fit will occur by chance alone. To identify such individuals, two separate goodness-of-fit tests were conducted for each student using their observed and expected PRC data on each of the parallel substrata. This yielded two chi-square model fit statistics for each student--one for each of the two sets of substrata. Assuming that the two chi-square values were independent, reliable person-non-fit would be indicated by identifying persons with significant ($p < .05$) chi-square values for each of the substrata tests of independence; the probability of observing such a result by chance alone would be .05 × .05, or .0025.

PRCs and ability level. If the responses of most persons fit the ICC model, the observed PRC should be a function of ability level (θ), just as the expected PRC is a function of ability level. To investigate this possibility, a variation of the D^2 reliability analysis was used. Based on observed PRC data within substrata, students were first matched on ability level ($\hat{\theta}$) before the between-persons D^2 measures were computed. These mean $\hat{\theta}$-matched between-persons D^2 values were then compared to the within-persons D^2 values, on the hypothesis that there should be little difference between

these means (and considerably less difference than when persons were matched without regard to $\hat{\theta}$) if observed PRCs were primarily a function of ability level.

Correlates of Observed PRCs

Pearson product-moment correlations were computed among scores on the three psychological reactions scales, the within-persons D^2, and the overall person-fit chi-square. Assuming the validity of the psychological reactions scales, it would be expected that both the D^2 and chi-square values would correlate positively with Perceived Test Difficulty and Test-Taking Anxiety. Chi-square and D^2 values were also correlated with ability estimates ($\hat{\theta}$).

Results

Test Characteristics

Table 1 shows the means, standard deviations, and range of item difficulties (*b*) and proportion-correct scores (*p*) in each of the nine strata, and the values of Cronbach's alpha internal consistency coefficient for each of the 24-item strata. The strata contained items of steadily increasing difficulty: Stratum 1 contained the easiest items and Stratum 9 contained the most difficult items. This distribution of items was mirrored in the proportion-correct data for each stratum. The average proportion correct decreased as difficulty level of items increased. An exception to this tendency occurred for Strata 8 and 9, in which average proportion correct was very similar. Although average proportion correct was related to the item difficulties in accordance with expectations, the data on the range of individual proportion-correct scores shows considerable variability in proportion correct within each of the nine strata. The largest range of proportion correct was in Stratum 4 where at least one student answered only .04 of the items correctly and the maximum observed proportion correct was 1.0. The smallest range of observed proportion correct was for Stratum 9, in which the minimum proportion-correct score was .04 and the maximum was .79. These data suggest a wide range of individual differences in the proportion-correct scores for each stratum and consequently the potential for individual differences in observed PRCs.

Table 1
Mean, Standard Deviation, and Range of Item Difficulties (*b*) and Proportion-Correct Scores (*p*), and Cronbach's Alpha Coefficient for Each of the Nine Strata

	Item Difficulties (*b*)				Proportion Correct (*p*)				
			Range				Range		
Stratum	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	Alpha
1	-2.40	.31	-2.97	-1.97	.866	.146	.130	1.000	.82
2	-1.55	.20	-1.93	-1.25	.794	.186	.210	1.000	.85
3	-1.01	.14	-1.24	-.77	.713	.216	.170	1.000	.86
4	-.56	.13	-.76	-.37	.615	.202	.040	1.000	.80
5	-.15	.11	-.36	.01	.545	.209	.080	1.000	.81
6	.26	.13	.06	.47	.481	.210	.040	.960	.80
7	.75	.19	.51	1.12	.416	.197	.080	.960	.78
8	1.32	.12	1.13	1.52	.330	.135	.040	.880	.54
9	1.98	.37	1.52	2.67	.334	.124	.040	.790	.44

Table 1 shows that the alpha internal consistency coefficients for Strata 1 through 7 were fairly high and quite similar, ranging from .78 to .86. Alpha coefficients for Strata 8 and 9 were lower--.54 and .44, respectively. The low alphas for Strata 8 and 9 were likely due to large amounts of random guessing for most students as the average porportion of correct responses of .33 for the two strata approached the theoretical expectation of .20 for the five-alternative multiple-choice items.

Table 2
Means and Standard Deviations of Item Difficulties (b) and Proportion-Correct Scores (p) in Each of the Nine Pairs of Parallel Substrata (A, B)

	Item Difficulties (b)				Proportion Correct (p)			
	Substratum				Substratum			
	A		B		A		B	
Stratum	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	-2.320	.296	-2.475	.317	.850	.156	.880	.166
2	-1.606	.240	-1.497	.153	.753	.202	.832	.198
3	-1.017	.161	-.993	.126	.713	.224	.711	.239
4	-.549	.137	-.572	.130	.622	.222	.606	.218
5	-.123	.084	-.180	.120	.545	.218	.543	.244
6	.296	.129	.219	.134	.460	.214	.501	.246
7	.762	.208	.740	.177	.401	.202	.429	.234
8	1.290	.113	1.354	.126	.341	.163	.317	.157
9	2.043	.411	1.910	.334	.295	.150	.370	.160

Table 2 provides data on each of the nine pairs of parallel substrata of 12 items each, including the means and standard deviations of item difficulties (b) and proportion-correct scores (p). Proportion-correct scores for each of the 151 students on each of the 18 substrata are in Appendix Table C, along with total proportion correct and the estimated ability level for each student. As Table 2 indicates, the substrata contained parallel items in the sense of similar means and standard deviations of difficulties. The smallest difference in mean difficulty was b=.002 for Stratum 3; the largest difference was b=.155 for Stratum 1, with a mean difference of .07. The proportion correct obtained by the students on the substrata were also fairly equal in mean and standard deviation. The smallest difference in mean proportion correct for the paired substrata was p=.002 for Stratum 3 and Stratum 5; the largest difference in mean observed proportion correct for the paired substrata was .075 (Stratum 2), indicating a high degree of similarity in mean proportion correct for the substrata.

Table 3 shows the estimated alpha coefficients for the 12-item substrata and the parallel forms correlations obtained by correlating proportion-correct scores for the 151 students on each of the nine pairs of substrata. The estimated 12-item alphas were obtained using the Spearman-Brown formula from the 24-item alphas for the strata shown in Table 1; these values were used in correcting for attenuation the parallel forms correlations. As Table 3 shows, the uncorrected parallel forms correlations between pairs of substrata ranged from .63 to .74 for the first seven strata; for the two most difficult strata the correlations were .42 and .28. These correlations were fairly substantial,

considering the low internal consistency reliabilities for the two most difficult strata. Using the Spearman-Brown formula to correct the parallel forms correlations based on two 12-item tests to the 24-item length of the strata, the average corrected correlation between the pairs of Substrata 1 through 7 was slightly above .80. For the most difficult two strata, the corrected correlations were .59 and .44.

Table 3
Estimated Alpha Coefficients for 12-Item Substrata, and Parallel Forms Correlation of Proportion-Correct Scores--Uncorrected, Corrected by Spearman-Brown Formula, and Corrected for Attenuation--on Each of the Nine Pairs of Parallel Substrata

		Parallel Forms Correlation		
Stratum	Estimated 12-Item Alpha	Uncorrected	Spearman-Brown Corrected	Attenuation Corrected
1	.69	.64	.78	.93
2	.74	.74	.85	1.00
3	.75	.73	.84	.97
4	.67	.70	.82	1.04
5	.68	.64	.78	.94
6	.67	.66	.80	.99
7	.64	.63	.77	.98
8	.40	.42	.59	1.00
9	.28	.28	.44	1.00

To determine whether scores on the paired substrata correlated as highly as possible, given the reliabilities of the substrata, the estimated 12-item alphas for the substrata were used along with the uncorrected parallel forms correlation to estimate the correlation between proportion-correct scores on the paired substrata, assuming that the substrata had been perfectly reliable. These attenuation-corrected correlations are shown as the last column in Table 3. As the data show, attenuation-corrected correlations were .97 or above for seven of the nine strata; for Strata 1 and 5, these correlations were .93 and .94, respectively. These data indicate that the paired substrata scores were as parallel as possible, given their estimated internal consistencies.

Reliability of Observed PRCs

Within- and between-persons D^2 indices. Table 4 shows summary statistics for the within-persons D^2 on the parallel substrata and the between-persons D^2 using randomly paired individuals. The within-persons D^2 mean of .28, with a standard deviation of .15 and range of .02 to .86, were all relatively small. These data indicate that for the within-persons D^2, the average difference in proportion correct on the paired substrata was about p=.18. By comparison, the between-persons D^2 mean was .75, with a standard deviation of .66 and a range of .07 to 4.09. Thus, the average difference in proportion correct between randomly paired individuals was about p=.29.

Table 4
Mean, Standard Deviation, and Range of Within- and Between-Persons D^2 Indices, and Results of t Tests Comparing the Mean Within-Persons D^2 with Each Between-Persons D^2 Index

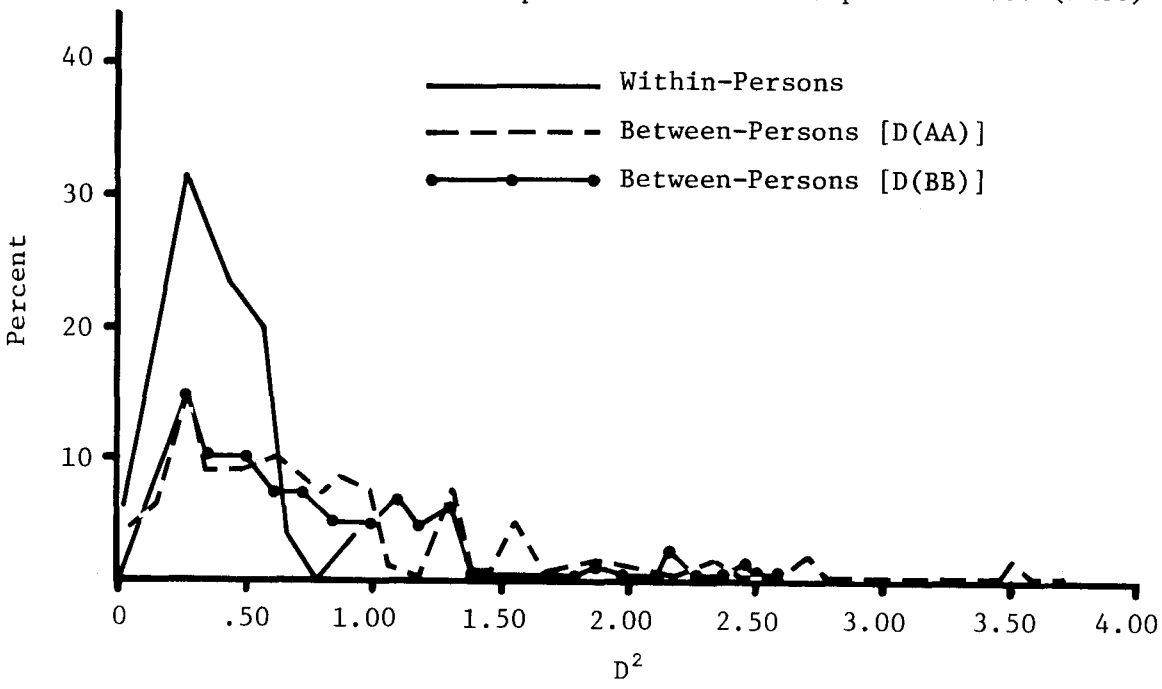
D^2 Index	N	Mean	SD	Range Min	Range Max	t	p*
Within-Persons	150	.28	.15	.02	.86		
Between-Persons							
D(AA)	75	.70	.64	.08	3.92	7.64	<.001
D(BB)	75	.78	.68	.06	4.32	8.62	<.001
D(AB)	75	.76	.69	.11	4.50	8.14	<.001
D(BA)	75	.76	.62	.04	3.60	9.06	<.001

*Probability of error in rejecting null hypothesis of no difference in group means.

The smaller within-persons D^2 demonstrates greater split-half profile similarity within persons than between pairs of randomly selected persons, irrespective of which split-half test was used for the between-persons comparisons. The t-test statistics in Table 4 demonstrate this sizable difference between the two types of profile comparison. Although the t-test assumption of independent groups was violated in these data, the mean differences in each case were substantial enough to support the conclusion that the PRCs are reliable.

Figure 4 provides further data on the distribution of the D^2 indices in terms of the relative frequency distributions of the within-persons reliability D^2 and two of the between-persons D^2 indices [D(AA) and D(BB)]. As Figure

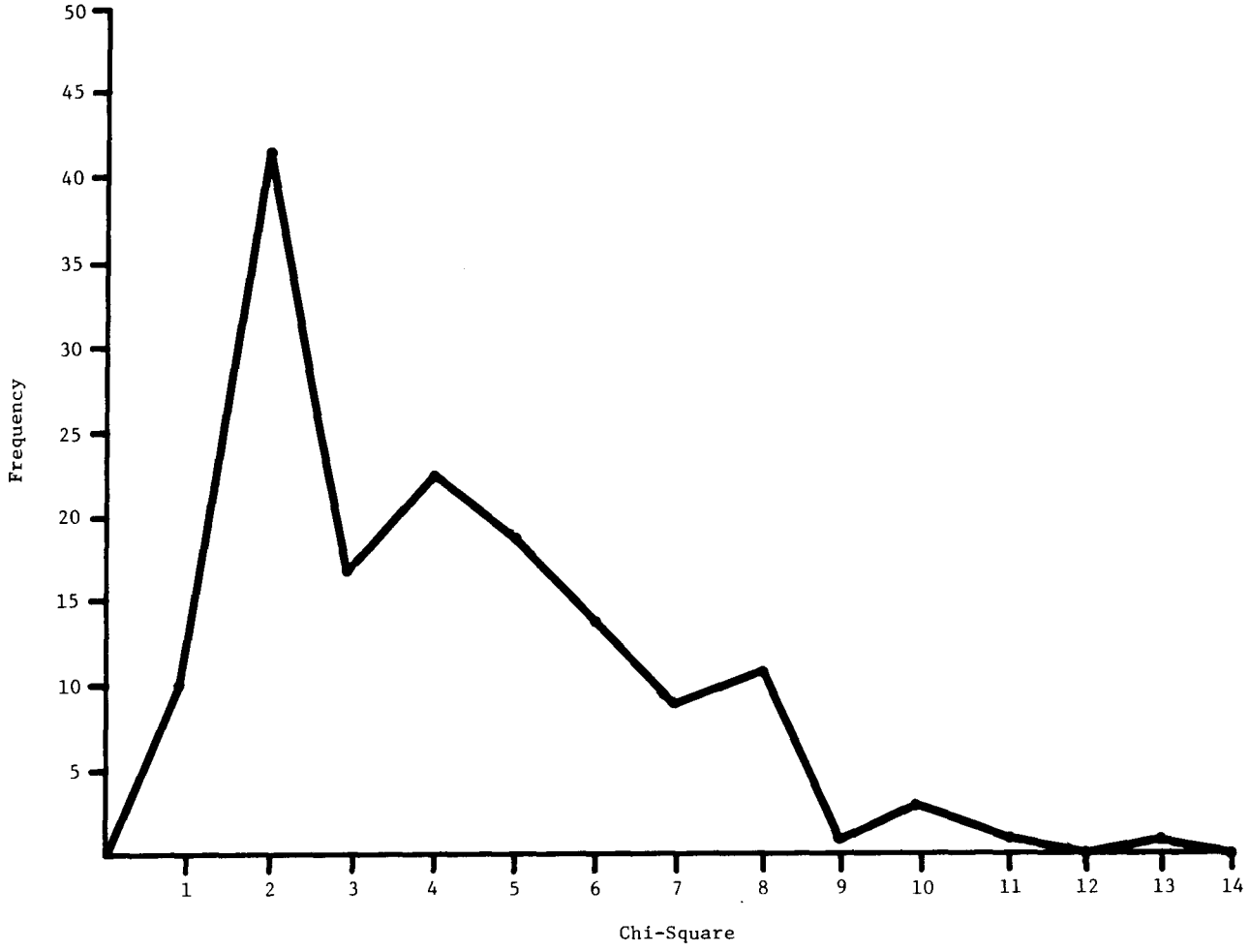
Figure 4
Relative Frequency Distributions of Within- and Between-Persons D^2 Indices for Observed Split-Half Person Response Curves (PRCs)



4 shows, there was little overlap between the two distributions. Virtually all of the within-persons D^2 values were below .75; and the distribution was highly peaked with a modal value very close to zero, indicating that the observed PRCs from the split substrata were very similar for most of the 150 students. By contrast, although the mode of between-persons D^2 indices was similar to that of the within-persons D^2, the relative frequency associated with that mode was considerably less than that of the within-persons distribution, and the distributions of the between-persons data was considerably less peaked.

Chi-square tests of independence. Results of the chi-square test of independence, based on a 2 × 9 contingency table with number-correct scores on each of the nine pairs of parallel substrata for each student, are shown in Figure 5. The minimum value of chi-square was .14 and the maximum was 12.94; mean of the distribution was 3.67, with a standard deviation of 2.34. A chi-square value of 15.51 is statistically significant at the p=.05 level with 8 degrees of freedom (from the 2 × 9 contingency table). Since all of the individual chi-square values were less than 15.51, the data show that the two split-pool observed PRCs for all students were not significantly different from each other, further supporting the D^2 data which indicated that the observed PRCs obtained from these data were reliable.

Figure 5
Frequency Polygon of Rounded Intra-Individual
Chi-Square Test of Independence Values Between Person
Response Curves (PRCs) for the Nine Pairs of Parallel Substrata

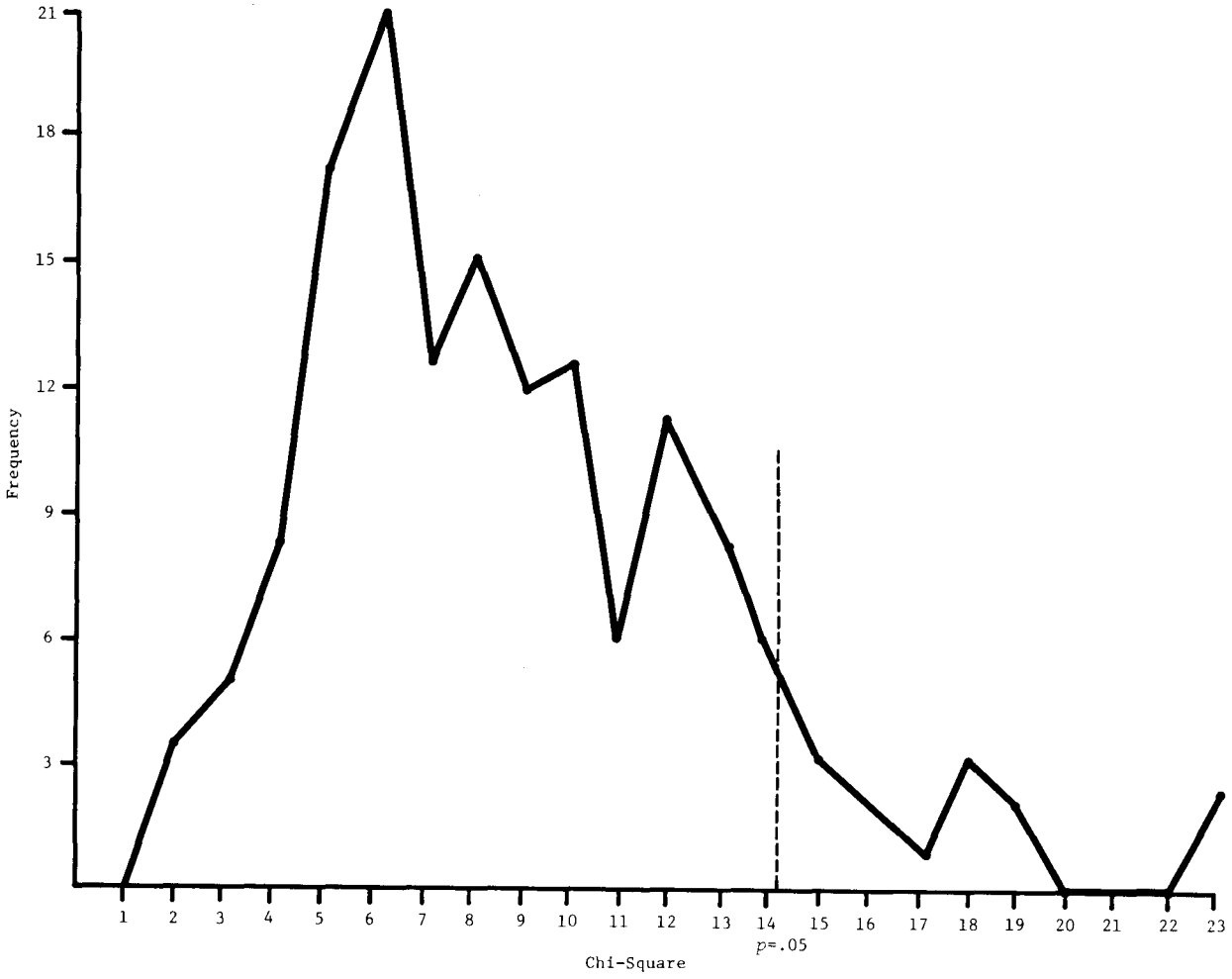


PRCs and Person-Fit

The frequency distribution of individual chi-square values reflecting the fit of observed PRCs to the PRCs expected from the three-parameter ICC model is shown in Figure 6. The lowest chi-square value obtained was 1.88 and the highest was 23.17. Mean chi-square was 8.76, with a standard deviation of 4.14; modal value was about 6.0.

Since ability estimates used in calculating the theoretically expected proportion-correct scores were taken from the data being analyzed, an extra degree of freedom was subtracted to determine the significance of the chi-square values. Thus, with 7 degrees of freedom, a chi-square value of 14.07 is significant at the .05 level. The group mean chi-square was well below this value, which would suggest that the three-parameter logistic ICC model served as a fairly good predictor of test response behavior for the majority of this group of students. Of 151 students, 8 would be expected to have significant chi-square values by chance alone at the .05 level; in this group, 15 students had chi-square values greater than 14.07.

Figure 6
Frequency Distribution of Intra-Individual Chi-Square Values for Goodness of Fit Between Observed and Expected Person Response Curves (PRCs)



To identify persons reliably deviating from the model, the chi-square person-fit statistics were recomputed for each student separately on the two sets of substrata. The joint distribution of chi-square values for the 151 students is shown in Figure 7, with the .05 significance level indicated by the dashed horizontal and vertical lines. Persons in the upper right-hand quadrant were identified as those deviating significantly from the expected values, with p=.0025. As Figure 7 shows, six students had significant chi-square values for both pairs of substrata and were thus placed in the upper right-hand quadrant. Of these six, four were also significantly non-fitting on the overall chi-square goodness-of-fit test. These four are indicated in Figure 7 by their subject numbers, and their PRCs (both observed and expected) are in Figure 9. Persons 83, 111, 138, and 117 might be hypothesized to have reliably non-fitting PRCs. Of the 15 students whose overall chi-square values were statistically significant, those not included in the upper right-hand quadrant may be hypothesized to be non-fitting only by chance.

Figure 7
Joint Distribution of Intra-Individual Chi-Square Values for Goodness of Fit for Odd- and Even-Numbered Substrata

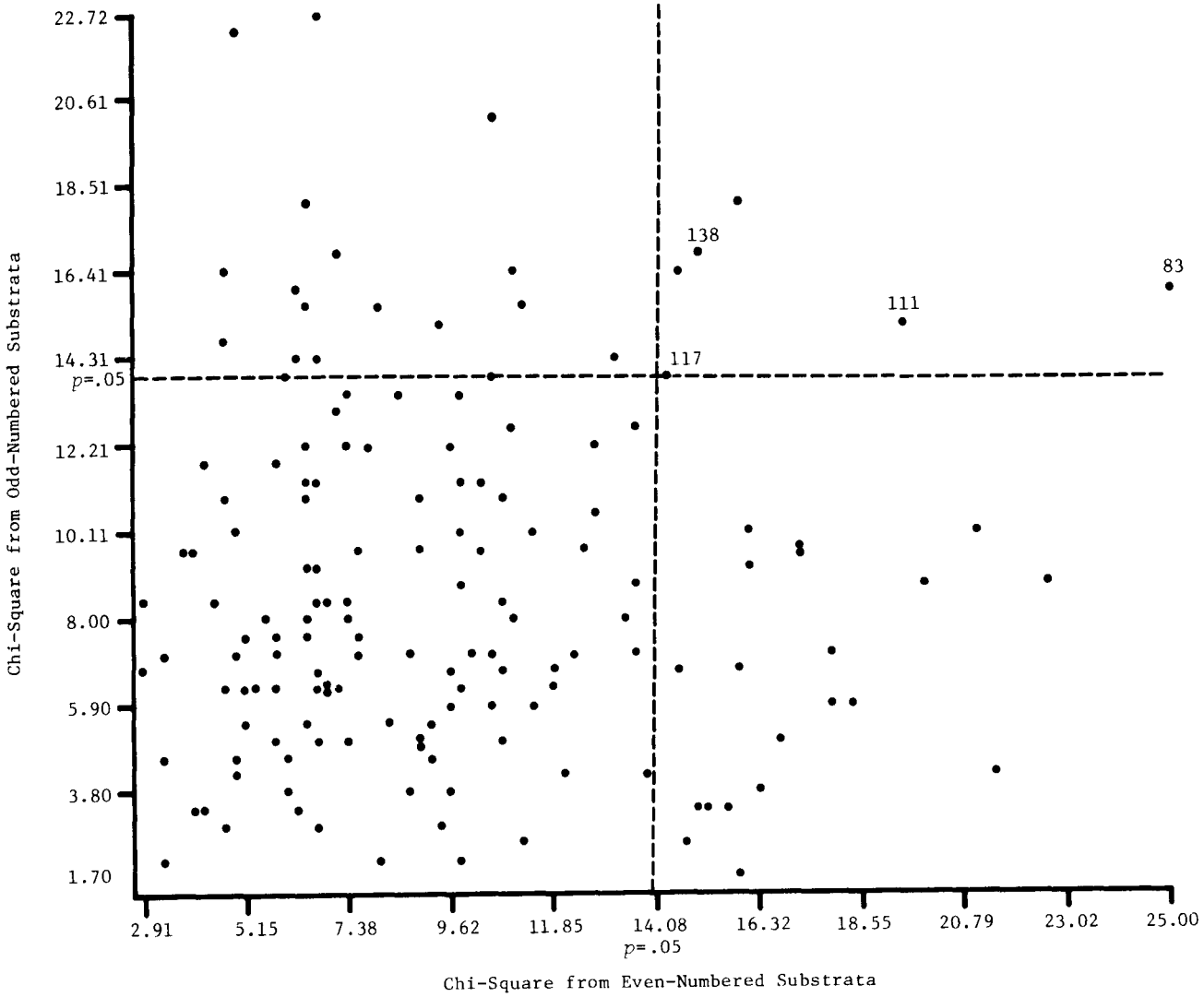


Figure 8 shows observed and expected PRCs for students with low overall chi-square person-fit values. Person 128 (Figure 8a) obtained the lowest chi-square value among the 151 students tested. As Figure 8a shows, the observed PRC for Person 128 (solid line) was quite close to the expected PRC (dashed line) for each of the nine strata. Figures 8b through 8d show expected and observed PRCs for three other students for whom model-fit was quite good, as indicated by the low chi-square values, although as expected, some minor deviations from model-fit appeared (e.g., Figure 8d) as chi-square values increased.

Figure 9 shows PRC person-fit results for four of the persons identified in Figure 7 as not reliably fitting the ICC model; these data are based on their total PRCs. The ways in which these four students' response curves deviated from their expected curves differed widely. Person 111 (Figure 9a) seems to have been careless with easier items, as indicated by a proportion correct of .75 on items in Stratum 1, and then to have been fortunate in guessing on some of the more difficult ones (p=.50 on Stratum 7). On the other hand this may be the type of profile to be expected from a person with an unusual educational history, such as an international student with a specialized knowledge of English. Person 117 (Figure 9b) and Person 138 (Figure 9d) seemed to have done much better on difficult items than was predicted by the model; these students might be sophisticated at guessing or high in "testwiseness." Person 83 (Figure 9c) seems to have exhibited carelessness on the easier items (Stratum 1) but more effort (with perhaps some good guesses) on the more difficult items in Strata 6 through 8.

Although these figures demonstrate lack of fit of individuals to the model-based predictions, they do not by themselves point to clear interpretations. However, they do illustrate some of the different ways in which significant deviations in test data can occur. This demonstrates the need for methods of assessing and interpreting the many ways in which non-fitting PRCs may occur.

PRCs and Ability Level

Additional data supporting the overall fit of persons to the three-parameter ICC model are shown in Table 5. Table 5 summarizes the distributions of

Table 5
Mean, Standard Deviation, and Range of Within-Persons D^2 and $\hat{\theta}$-Matched Between-Persons D^2, and Results of t Tests Comparing the Mean Within-Persons D^2 with Each Between-Persons D^2 Index

D^2 Index	N	Mean	SD	Range Min	Range Max	t	p*
Within-Persons	150	.28	.15	.02	.86		
Between-Persons							
D(AA)	75	.25	.11	.03	.48	1.50	<.20
D(BB)	75	.26	.13	.05	.74	1.00	<.50
D(AB)	75	.29	.14	.05	.79	0.50	<.80
D(BA)	75	.28	.12	.07	.64	0.00	1.00

*Probability of error in rejecting null hypothesis of no difference in group means (two-tailed test).

Figure 8
Observed and Expected Person Response Curves (PRCs) for Four Persons Whose Responses Reliably Fit the Three-Parameter ICC Model

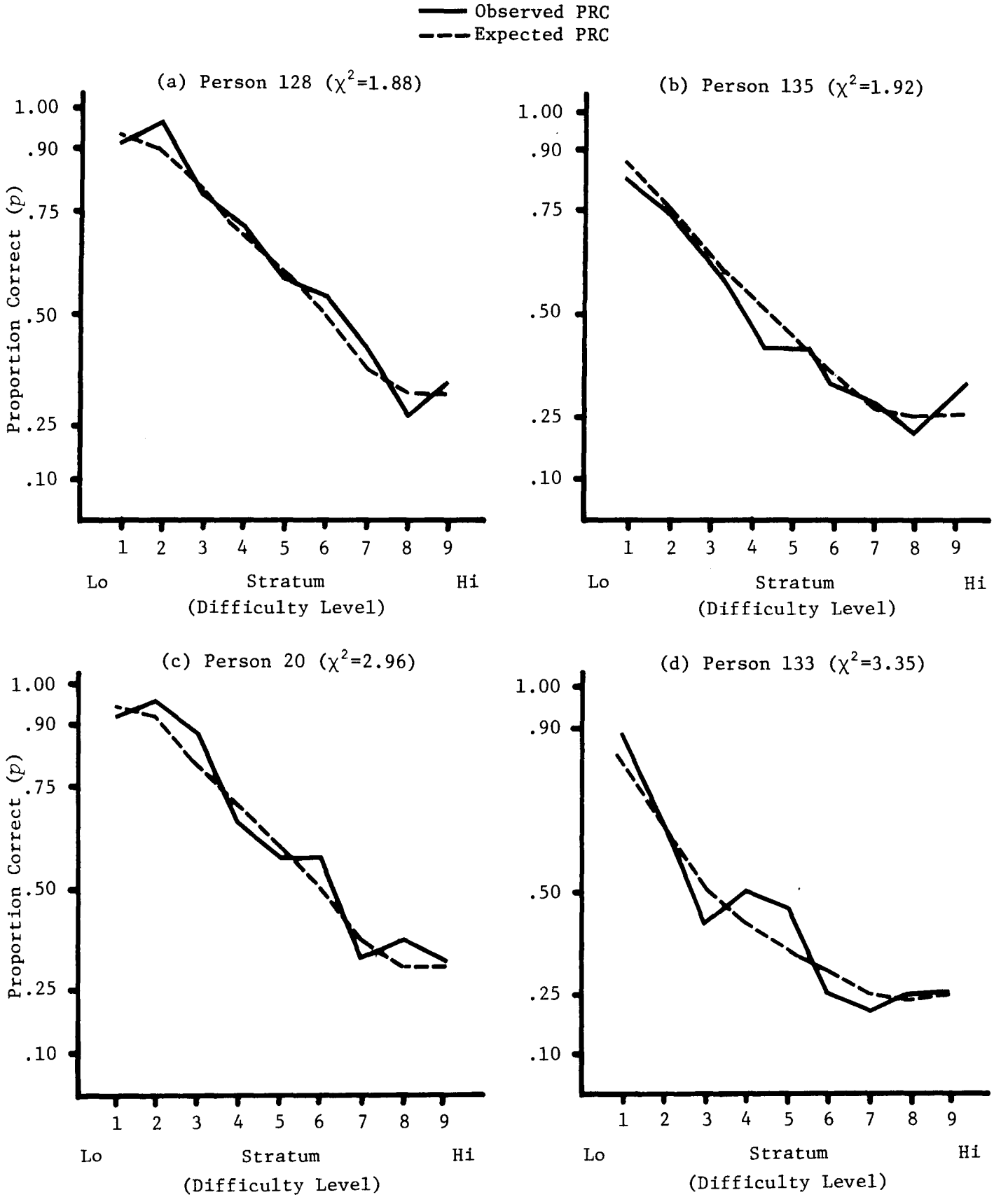
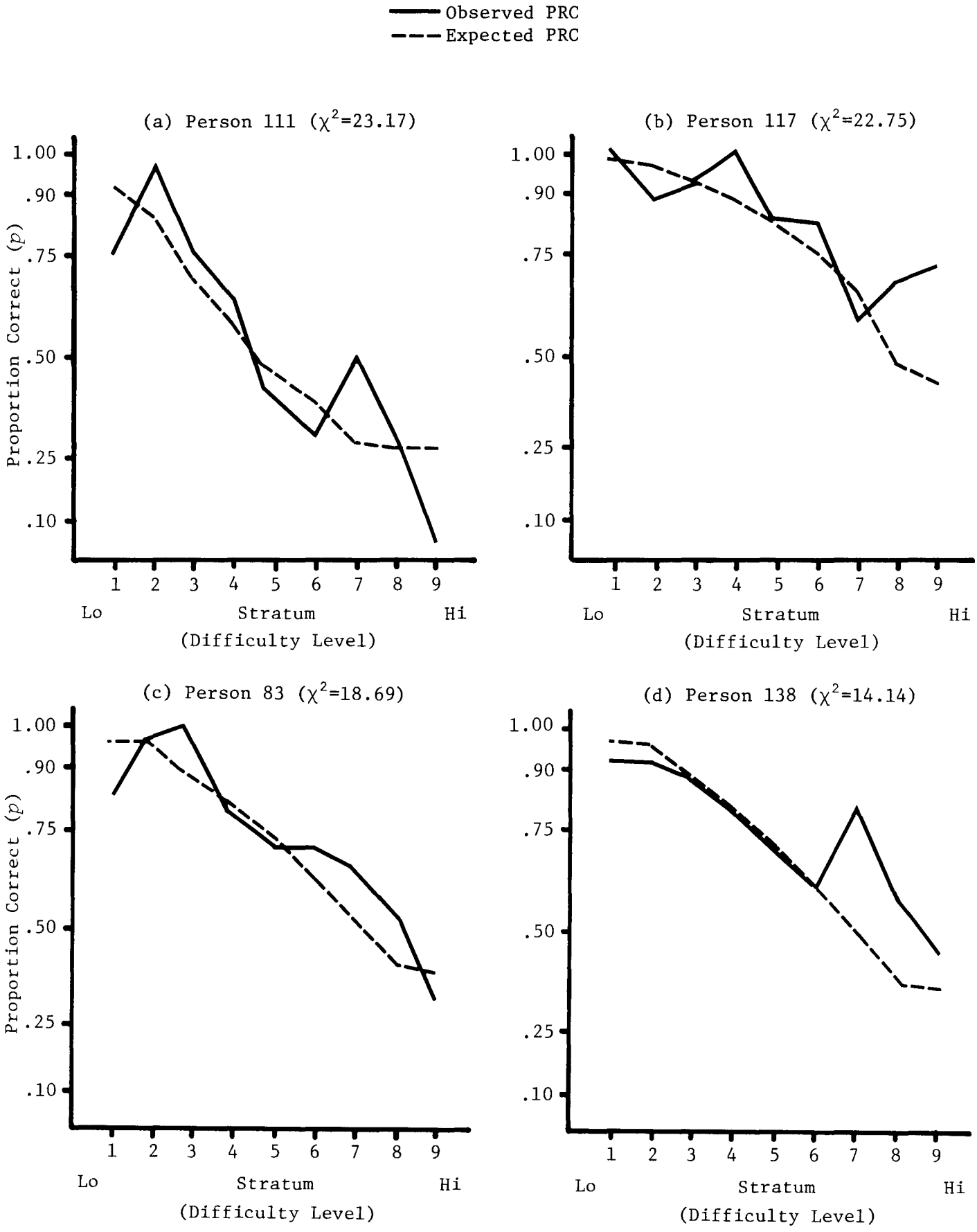


Figure 9
Observed and Expected Person Response Curves (PRCs) for Four Persons Whose Responses Did Not Reliably Fit the Three-Parameter ICC Model



between-persons D^2 data on parallel substrata when students were matched as closely as possible for $\hat{\theta}$ values before the substrata D^2 indices were calculated. As Table 5 shows, none of the mean between-persons D^2 indices was significantly different from the within-persons D^2; in three of the four cases the mean between-persons D^2 was slightly lower than the mean within-persons D^2. In addition, the standard deviations and ranges of the two kinds of D^2 indices were very similar. Thus, the data in Table 5 show that observed PRCs for this group of students were highly dependent upon their ability levels, further supporting the fit of these individuals to the three-parameter ICC model.

Correlates of Observed PRCs

Table 6 shows intercorrelations of the within-persons D^2 PRC reliability index; the PRC person-fit chi-square value for each person; ability estimates ($\hat{\theta}$); and the Perceived Test Difficulty, Test-Taking Anxiety, and Test-Taking Motivation scale scores. The D^2 reliability indices correlated significantly (r=-.24) with ability, indicating a tendency for lower ability students to have more unreliable PRCs. D^2 also correlated significantly positively with both Perceived Test Difficulty and Test-Taking Anxiety scale scores; the correlation with Perceived Test Difficulty scores probably reflected the high negative correlation (r=-.70) between ability level and perceived difficulty of the test items. The correlation of r=.18 with Test-Taking Anxiety suggests a tendency for students with higher test-taking anxiety to have less reliable PRCs. None of the correlations of the chi-square person-fit index were statistically significant. Further analysis of the relationship of the chi-square person-fit indices by analysis of variance indicated no nonlinear relationships between the chi-square index and the Perceived Test Difficulty, Test-Taking Anxiety, and Test-Taking Motivation scale scores.

Table 6
Intercorrelations of Ability Estimates ($\hat{\theta}$), Psychological Reactions Scales, and PRC Within-Persons D^2 and Person-Fit Chi-Square Indices

	Ability	Perceived Test Difficulty	Test-Taking Anxiety	Test-Taking Motivation	Within-Persons D^2
Ability					
Perceived Test Difficulty	-.70**				
Test-Taking Anxiety	-.16*	.28**			
Test-Taking Motivation	.37**	-.35**	.11**		
Within-Persons D^2	-.24**	.18**	.18**	.03	
Person-Fit Chi-Square	-.06	-.05	.07	.04	-.04

*Significant at p<.05.
**Significant at p<.01.

These results are consistent with the previously reported findings that the three-parameter logistic model seemed to predict quite well the test performance of the majority of the students in this sample. Since only a few of the students deviated significantly and reliably from the predictions from the model, it would be impossible to find strong relationships between the goodness-of-fit results and other variables. Furthermore, as was illustrated in Figure 9, there are many possible ways of deviating from the model and, consequently, there may be many correlates of such deviations.

Conclusions and Directions for Future Research

The feasibility of the person response curve (PRC) approach to investigating the fit of persons to the three-parameter ICC model was explored in this study. To operationalize the PRC it was necessary to subdivide ability test items into separate strata of varying difficulty levels. For the vocabulary test used in this study, strata possessed sufficient internal consistency and parallel forms reliability to justify their use, although the more difficult strata were much less reliable than the easier strata.

Conclusions

The PRCs proved to be highly reliable. The D^2 analyses indicated not only that intra-individual profiles were more similar than profiles between randomly selected persons but also that profiles between people of similar ability level were also very similar. As additional evidence of profile reliability, chi-square tests for independence between profiles of parallel forms for each individual were nonsignificant for all 151 students. The high correlation of r=.82 (p<.001) between the intra-individual parallel forms chi-squares and the D^2 suggests that the chi-square test may be sufficient in future studies, since it also provides a more ready means of assessing statistical significance.

The results of the D^2 statistics between individuals matched on ability level were interesting, since they illustrated close profile similarity between different persons of similar ability level. This suggests that for the majority of this sample, PRCs were predictable as a function of ability level. A more complete test of this hypothesis was conducted with a chi-square goodness-of-fit test between observed proportion-correct scores on each of nine strata and expected proportion-correct scores predicted by the three-parameter logistic model. The nonsignificant group mean suggests that the model was a reasonable way of describing students' test response behavior.

At the .05 level, eight students were expected to have significant chi-square goodness-of-fit values for observed and expected PRCs. Fifteen students had significant chi-square values, leaving somewhat in question whether these students deviated from the model because of chance or interaction with another dimension. One method of investigating this question was to calculate separately the goodness of fit of each student's observed and expected PRCs on the odd-numbered substrata and on the even-numbered substrata. Of the 15 students with significant chi-square values on the overall nine-strata goodness-of-fit test, four had significant chi-square values on both substrata goodness-of-fit tests. These four students were identified as reliably deviating from the ICC model predictions. The nature of this lack of fit, however, would best be investigated in a future study with an experimental design that included interactions with additional dimensions other than the ability being measured.

Having demonstrated the goodness of fit of observed PRCs with model-predicted PRCs, and with no firm evidence to suggest that significant results for a majority of the students were due to anything other than chance, the nonsignificant results for the relationship of the goodness-of-fit chi-square variable with nontest variables seems to follow. Scores on the psychological reactions scales correlated with each other and with ability estimates in expected ways but did not correlate significantly with the overall chi-square

variable. These results substantiated the fit of the model to observed student test-response behavior. The psychological reactions scales could be used in a future study of non-fit in which these psychological states could be experimentally induced.

The results of this study demonstrate that the PRC can be useful in studying the fit of individuals to ICC models by testing the fit of the observed PRC to the theoretically expected PRC. Although the three-parameter ICC model was used here, the method can be used with the two-parameter or one-parameter logistic (Rasch) model or with any of the normal ogive ICC models. The data also demonstrated that the three-parameter ICC model adequately accounted for the test response behavior of the vast majority of the students studied. More research is, of course, necessary to further explore the use of the PRC in examining model-fit in test behavior.

Directions for Future Research

Guessing and "testwiseness" are variables which are unrelated to abilities but may affect ability test scores. To determine whether these variables can be detected by PRCs or PRC-fit to theoretical predictions, a useful experiment would be to administer a multiple-choice ability test along with testwiseness and guessing scales to groups of students. One subgroup in the experimental design should be an experimental group trained in testwiseness and/or in guessing skills. The effects of testwiseness or guessing would be studied by analysis of the chi-square goodness-of-fit statistics comparing the expected and observed PRCs for the experimental and control groups. Special attention should be given to chi-square values on the most difficult items in the ability test rather than overall chi-squares, since it is on these items that the experimental effect is likely to be observed.

Cultural bias is another dimension which may differentially affect ability test performance (e.g., Church, Pine, & Weiss, 1978; Martin, Pine, & Weiss, 1978; Pine & Weiss, 1978). One approach to testing for the existence of such bias by use of PRCs would be to compare the goodness of fit of observed and expected PRCs for a control group of white middle-class testees and a group of testees who would be hypothesized to have uneven educational development by white middle-class American standards. This latter group might involve international students with a specialized knowledge of the English language or some American minority group persons. It would be expected that the PRCs would show greater deviation from the model predictions for the latter group, particularly in terms of deviations from the unidimensionality required by the ICC model.

Carelessness and nervousness are two other dimensions which may contribute to unexpected performance on ability tests and which may be detected by PRC analysis. To study the effect of these dimensions on person-fit, an ability test could be administered to three groups of randomly selected individuals from the same population. A low-motivation-possibly-careless control group would be given minimal information about the test. Treatment Group 1 would be told that the test results did not matter and that the experimenter just needed to fill his/her quota of subjects. Treatment Group 2 would be told that the test is an important determiner of whether or not they would be able to complete college or to succeed in some occupation; this would be considered the high-anxiety group. The experimentally induced states should be verified with

improved versions of the psychological reactions scales for motivation and anxiety used in this report. Values comparing chi-square observed versus expected PRCs would be compared, with special attention to the PRC-fit data on the easier strata for the low-motivation group and on the more difficult strata for the high-anxiety group. This would give information on possible psychological correlates of fit on a stratum-by-stratum basis. Data of this type might be used, for example, to investigate the operation of the Yerkes-Dodson Law (Taylor & Spence, 1958; Yerkes & Dodson, 1908) in ability test data; PRC-fit data would support this hypothesis if high-anxiety testees perform better than expected on easy test items and more poorly than expected on the more difficult test items.

Further investigation of the measurement properties of observed versus expected chi-square goodness-of-fit statistics for assessing non-fit of persons is also of importance. Monte carlo simulations should be run in order to determine the null distribution of the chi-square values. These should be repeated at a number of theta levels to determine whether goodness-of-fit distributions differed as a function of ability level.

Finally, since the research literature on methods for assessing non-fitting profiles has begun to branch in several different directions, it would be informative and useful to compare the efficacy of several different methods using the same data base. The one-, two-, and three-parameter ICC models could each be used in computing ability estimates so that non-fit measures based on these different models could be used. This would best be done in simulation, with non-fitting data experimentally induced so that the different methods of evaluating model-fit could be compared on their degree of "hits" and "misses."

These are only a few of many research possibilities in investigating the properties and the diagnostic utility of PRCs. A closer look at these properties of the PRC test performance profiles and their use in determining person-fit may provide important information on selected individuals and improve the validity of ability tests for individual prediction and diagnosis.

References

Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1979. (NTIS No. AD A067752)

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828)

Betz, N., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (NTIS No. AD A027170)

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977. (NTIS No. AD A046062)

Church, A. T., Pine, S. M., & Weiss, D. J. A comparison of levels and dimensions of performance in black and white groups on tests of vocabulary, mathematics, and spatial ability (Research Report 78-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1978. (NTIS No. AD A062797)

Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. Psychological Bulletin, 1953, 50, 456-473.

Hambleton, R., & Cook, L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

Kingsbury, G. G., & Weiss, D. J. Relationships among achievement level estimates from three item characteristic curve scoring methods (Research Report 79-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1979. (NTIS No. AD A069815) (a)

Kingsbury, G. G., & Weiss, D. J. An adaptive testing strategy for mastery decisions (Research Report 79-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1979. (b)

Levine, M. Psychometric models and appropriateness measurement. In D. J. Weiss (Ed.), Proceedings of the 1979 conference on computerized adaptive testing, in preparation.

Levine, M., & Rubin, D. Measuring the appropriateness of multiple-choice test scores (Research Bulletin 76-31). Princeton, NJ: Educational Testing Service, December 1976.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lord, F. M., & Novick, M. Statistical theories of mental test scores. Reading, MA: Addison & Wesley, 1968.

Lumsden, J. Person reliability. Applied Psychological Measurement, 1977, 1, 477-482.

Lumsden, J. Tests are perfectly reliable. British Journal of Mathematical and Statistical Psychology, 1978, 31, 19-26.

Martin, J. T., Pine, S. M., & Weiss, D. J. An item bias investigation of a standardized aptitude test (Research Report 78-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1978. (NTIS No. AD A064352)

McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement (Research Report 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1974. (NTIS No. AD 781894)

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976. (NTIS No. AD A022964)

Mead, R. J. Using the Rasch model to identify person-based measurement disturbances. In D. J. Weiss (Ed.), Proceedings of the 1979 conference on computerized adaptive testing, in preparation.

Mosier, C. I. Psychophysics and mental test theory: Fundamental postulates and elementary theorems. Psychological Review, 1940, 47, 355-366.

Mosier, C. I. Psychophysics and mental test theory II: The constant process. Psychological Review, 1942, 48, 235-249.

Pine, S. M., & Weiss, D. J. A comparison of the fairness of adaptive and conventional testing strategies (Research Report 78-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, August 1978. (NTIS No. AD A059436)

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Schmidt, F., & Urry, V. W. Item parameterization procedures for the future. In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing (U.S. Civil Service Commission, Personnel Research and Development Center, PS-75-6). Washington, DC: U.S. Government Printing Office, 1976. (Superintendent of Documents Stock No. 006-00940-9)

Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. Journal of Experimental Psychology, 1958, 57, 55-60.

Vale, C. D., & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1975. (NTIS No. AD A018758)

Vale, C. D., & Weiss, D. J. A comparison of information functions of multiple-choice and free-response vocabulary items (Research Report 77-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1977.

Wainer, H., & Wright, B. D. Robust estimation of ability in the Rasch model. In D. J. Weiss (Ed.), Proceedings of the 1979 conference on computerized adaptive testing, in preparation.

Weiss, D. J. Canonical correlation analysis in counseling psychology research. Journal of Counseling Psychology, 1972, 19, 241-252.

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376)

Weiss, D. J. (Ed.). Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975. (NTIS No. AD A018675)

Wood, R. L., Wingersky, M. S., & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (Research Memorandum 76-6). Princeton, NJ: Educational Testing Service, 1976. (modified January 1978)

Wright, B. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-115.

Wright, B., & Mead, R. J. The use of measurement models in the definition and application of social science variables (Technical Report DAHC19-76-G-011). Arlington, VA: U.S. Army Research Institute for the Behavioral Sciences, 1977.

Wright, B. D., & Stone, M. H. Best test design. Chicago: MESA Press, 1979.

Yerkes, R. M., & Dodson, J. D. The relation of strength to stimulus to rapidity of habit formation. Journal of Comparative Neurological Psychology, 1908, 18, 459-482.

Table A

Item Numbers, Discrimination (a) and Difficulty (b) Parameters, and Substratum Designation for Items in the Vocabulary Item Pool (c=.20 for All Items)

Item Number	a	b	Sub-stratum	Item Number	a	b	Sub-stratum	Item Number	a	b	Sub-stratum	Item Number	a	b	Sub-stratum
Stratum 9				Stratum 7, *cont'd.*.				Stratum 5, *cont'd.*				Stratum 3, *cont'd.*			
343	.264	2.671	A	523	1.210	.875	B	270	1.223	-.138	A	199	1.093	-1.093	B
580	.315	2.639	B	302	.845	.846	B	143	1.036	-.153	A	108	.536	-1.155	A
243	.316	2.607	A	271	.886	.796	A	156	.841	-.166	A	86	.887	-1.189	A
228	2.935	2.411	A	139	.614	.794	B	643	.487	-.202	A	189	.757	-1.191	B
569	.400	2.402	A	324	.524	.772	A	211	.773	-.236	A	141	.478	-1.208	A
394	.297	2.398	B	267	.650	.770	A	37	.860	-.236	B	227	.812	-1.245	A
167	.416	2.155	A	289	.480	.691	B	157	.384	-.245	A				
533	.632	2.153	B	113	1.057	.678	B	390	.797	-.257	B	Stratum 2			
247	.647	2.063	A	340	1.921	.645	B	224	.679	-.257	B	232	.673	-1.251	B
577	.613	2.004	B	60	1.232	.643	B	221	.822	-.278	B	191	1.749	-1.257	A
374	.557	1.992	A	590	.538	.617	A	307	.699	-.325	B	88	.706	-1.332	B
531	.346	1.921	B	59	1.093	.601	B	128	1.074	-.355	B	186	1.067	-1.335	A
260	.709	1.820	B	372	.346	.559	A					127	1.075	-1.345	A
504	.635	1.808	A	593	.560	.551	B	Stratum 4				129	1.274	-1.352	B
603	.380	1.800	B	264	2.276	.549	A	535	.767	-.374	A	101	1.165	-1.395	A
616	.610	1.759	A	265	1.571	.546	A	58	.587	-.380	A	44	1.145	-1.412	B
119	.534	1.729	B	538	1.181	.518	A	203	.820	-.384	B	311	.746	-1.430	A
521	.753	1.696	B	266	2.120	.509	B	33	.800	-.390	B	190	1.818	-1.439	B
400	.929	1.682	B					332	.973	-.396	A	83	.875	-1.449	B
242	.524	1.574	B	Stratum 6				130	.949	-.439	A	214	.476	-1.488	B
610	.788	1.566	A	252	.420	.472	B	183	.728	-.452	B	13	1.888	-1.553	B
159	.768	1.559	A	633	.712	.470	A	588	.465	-.464	B	34	.830	-1.582	B
168	.913	1.553	B	301	1.376	.468	A	53	.637	-.478	B	84	1.701	-1.640	A
350	.317	1.525	A	519	.527	.440	A	222	.652	-.499	A	559	.616	-1.675	A
				377	.585	.393	B	142	.314	-.536	A	27	1.427	-1.675	B
Stratum 8				582	1.200	.351	A	123	.823	-.559	A	95	.563	-1.707	B
383	2.111	1.518	B	549	.433	.348	A	136	.317	-.562	B	96	1.129	-1.722	B
525	.570	1.509	B	551	.896	.336	B	293	.669	-.567	A	76	.618	-1.750	A
147	.825	1.469	A	116	.494	.334	B	287	.523	-.652	A	196	2.128	-1.791	A
253	2.321	1.443	B	50	.694	.321	A	117	.619	-.656	B	197	.253	-1.850	A
572	1.289	1.433	B	318	.526	.310	A	85	.934	-.670	B	125	1.236	-1.875	A
213	.429	1.430	A	272	1.960	.223	A	584	.758	-.677	A	262	.768	-1.928	A
368	.462	1.424	B	502	.730	.218	B	185	.682	-.684	B				
216	.668	1.397	B	52	.844	.205	A	109	1.109	-.701	B	Stratum 1			
217	1.249	1.384	A	54	.378	.204	B	515	1.084	-.708	B	22	1.200	-1.971	B
660	.829	1.369	B	622	.444	.201	A	239	.939	-.709	B	158	1.083	-1.996	A
291	1.641	1.354	A	599	1.634	.158	B	204	.876	-.742	A	106	.672	-2.009	A
333	.351	1.340	B	354	.327	.151	A	87	1.241	-.763	A	138	1.728	-2.023	A
397	.651	1.339	A	56	1.109	.135	B					31	.722	-2.141	B
586	1.536	1.309	A	161	1.384	.132	B	Stratum 3				63	.692	-2.144	A
268	.270	1.300	A	355	.506	.104	B	112	.614	-.775	A	202	.620	-2.172	B
259	.365	1.293	B	145	.791	.086	B	235	.664	-.776	A	206	1.105	-2.187	A
341	.634	1.282	A	209	.870	.067	A	36	1.644	-.789	B	184	.726	-2.193	A
581	1.256	1.207	A	444	.621	.059	B	546	.555	-.801	B	9	1.452	-2.240	B
306	1.317	1.204	B					615	.439	-.858	B	80	.859	-2.251	B
231	.874	1.186	B	Stratum 5				43	1.108	-.861	A	126	.956	-2.266	A
617	2.778	1.172	A	292	.610	.012	B	371	.444	-.916	B	602	.255	-2.285	A
164	.687	1.136	B	597	.624	-.000	A	194	1.790	-.959	A	68	1.014	-2.479	A
238	.758	1.130	A	382	.856	-.010	A	47	1.043	-.962	A	198	.801	-2.503	B
576	.427	1.128	A	205	.603	-.024	B	103	1.059	-.999	B	131	.604	-2.577	B
				207	.793	-.035	A	26	.364	-1.020	A	181	1.020	-2.584	B
Stratum 7				137	.499	-.056	A	285	.835	-1.022	A	151	.438	-2.651	A
516	.350	1.116	B	503	1.062	-.090	B	637	.877	-1.023	B	48	.266	-2.696	B
601	1.315	1.097	A	365	.877	-.105	A	40	1.236	-1.032	A	65	1.024	-2.711	B
215	.908	1.069	A	176	.415	-.106	B	51	1.432	-1.043	B	135	.425	-2.789	A
111	.822	.936	A	154	.872	-.124	B	241	.568	-1.054	B	121	.743	-2.820	A
375	.832	.934	B	218	.407	-1.25	B	173	.882	-1.062	B	17	.716	-2.891	B
526	1.169	.919	A	234	.650	-.132	A	322	.673	-1.091	B	201	.310	-2.966	B

Table B

Test Reaction Items Used for the Perceived Test Difficulty and Test-Taking Anxiety and Motivation Scales, and Scoring Weights for Each Response

Scale and Item	Scoring Weight
Perceived Test Difficulty	
1. How often did you feel that the questions in the test were too easy for you?	
(1) Always	1
(2) Frequently	2
(3) Sometimes	3
(4) Seldom	4
(5) Never	5
6. In relation to your vocabulary ability, how difficult was the test for you?	
(1) Much too difficult	-1
(2) Somewhat too difficult	-2
(3) Just about right	-3
(4) Somewhat too easy	-4
(5) Much too easy	-5
9. How often were you sure that your answers to the questions were correct?	
(1) Almost always	1
(2) More than half of the time	2
(3) About half of the time	3
(4) Less than half of the time	4
(5) Almost never	5
13. How often did you feel that the questions in the test were too hard for you?	
(1) Always	-1
(2) Frequently	-2
(3) Sometimes	-3
(4) Seldom	-4
(5) Never	-5
Test-Taking Anxiety	
2. How did you feel while taking the test?	
(1) Very tense	-1
(2) Somewhat tense	-2
(3) Neither tense nor relaxed	-3
(4) Somewhat relaxed	-4
(5) Very relaxed	-5
Test-Taking Anxiety, *cont'd.*	
4. During testing, did you worry about how well you would do?	
(1) Not at all	1
(2) Very little	2
(3) Somewhat	3
(4) Fairly much so	4
(5) Very much so	5
7. Did nervousness while taking the test prevent you from doing your best?	
(1) Definitely	-1
(2) Probably	-2
(3) Not sure	-3
(4) Probably not	-4
(5) Definitely not	-5
11. Were you nervous while taking the test?	
(1) Not at all	1
(2) Somewhat	2
(3) Very little	3
(4) Moderately so	4
(5) Very much so	5
Test-Taking Motivation	
3. Did you feel challenged to do as well as you could on the test?	
(1) Not at all	1
(2) Very little	2
(3) Somewhat	3
(4) Fairly much so	4
(5) Very much so	5
10. Did you care how well you did on the test?	
(1) I cared a lot	-1
(2) I cared some	-2
(3) I cared a little	-3
(4) I cared very little	-4
(5) I didn't care at all	-5
12. Do you think that you could have done better on the test if you had tried harder?	
(1) I definitely could have	1
(2) I probably could have	2
(3) I'm not sure	3
(4) I probably couldn't have	4
(5) I definitely couldn't have	5

Table C
Ability Estimate ($\hat{\theta}$), Total Proportion Correct (T), and Proportion Correct for Each Student on Each of the Substrata (A, B) of the Nine-Stratum Test

			Stratum and Substratum																	
			1		2		3		4		5		6		7		8		9	
Student	$\hat{\theta}$	T	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	-1.01	.47	.86	.88	.77	.76	.60	.60	.49	.50	.41	.42	.34	.35	.27	.27	.25	.25	.26	.26
2	-1.87	.35	.70	.73	.51	.48	.39	.39	.34	.34	.29	.30	.27	.27	.23	.23	.23	.23	.23	.23
3	-.11	.61	.94	.95	.92	.92	.82	.81	.71	.71	.60	.62	.50	.52	.37	.37	.31	.31	.31	.31
4	-1.50	.40	.78	.81	.63	.60	.47	.47	.39	.40	.33	.34	.29	.30	.24	.24	.24	.23	.24	.24
5	-.28	.59	.93	.94	.90	.90	.78	.78	.67	.67	.56	.58	.46	.48	.34	.34	.30	.30	.30	.30
6	-.75	.51	.89	.91	.83	.82	.67	.67	.55	.56	.46	.47	.38	.39	.28	.29	.27	.27	.27	.27
7	-1.50	.40	.78	.81	.63	.60	.47	.47	.39	.40	.33	.34	.29	.30	.24	.24	.24	.24	.24	.24
8	-1.70	.38	.74	.77	.56	.53	.43	.42	.36	.36	.31	.31	.28	.28	.23	.23	.23	.23	.24	.23
9	1.45	.84	.98	.98	.98	.99	.96	.96	.93	.93	.89	.90	.84	.85	.79	.80	.65	.62	.50	.53
10	-.68	.52	.90	.91	.85	.83	.69	.69	.57	.58	.47	.49	.39	.40	.29	.29	.27	.27	.27	.27
11	-.72	.52	.90	.91	.84	.82	.68	.67	.56	.57	.46	.48	.38	.39	.29	.29	.27	.27	.27	.27
12	-.95	.48	.87	.89	.79	.77	.61	.61	.51	.51	.41	.43	.35	.36	.27	.27	.26	.26	.26	.26
13	.44	.70	.96	.97	.96	.96	.89	.89	.81	.82	.73	.74	.64	.66	.50	.52	.38	.38	.36	.37
14	-.94	.48	.87	.89	.79	.77	.62	.61	.51	.52	.52	.42	.43	.35	.36	.27	.27	.26	.26	.26
15	.41	.69	.96	.97	.95	.95	.89	.89	.81	.81	.72	.74	.63	.65	.50	.51	.37	.37	.36	.37
16	-.71	.52	.90	.91	.84	.83	.68	.68	.56	.57	.46	.48	.38	.40	.29	.29	.27	.27	.27	.27
17	-.89	.49	.88	.89	.80	.78	.63	.63	.52	.53	.43	.44	.36	.37	.27	.27	.26	.26	.26	.26
18	-.50	.55	.92	.93	.87	.87	.73	.73	.61	.62	.51	.53	.42	.44	.31	.31	.28	.28	.29	.28
19	.87	.76	.97	.98	.97	.97	.93	.93	.88	.88	.81	.82	.74	.76	.64	.65	.47	.46	.42	.43
20	-.11	.61	.94	.95	.92	.92	.81	.81	.70	.71	.60	.62	.50	.52	.37	.37	.31	.31	.31	.31
21	-1.63	.38	.75	.78	.58	.56	.44	.44	.37	.37	.31	.32	.28	.29	.24	.24	.23	.23	.24	.24
22	-.98	.48	.87	.88	.78	.76	.61	.60	.50	.51	.41	.43	.34	.35	.27	.27	.25	.25	.26	.26
23	-.43	.56	.92	.93	.88	.88	.75	.75	.63	.64	.53	.54	.43	.45	.32	.32	.28	.29	.29	.29
24	.93	.77	.97	.98	.97	.98	.93	.93	.88	.89	.82	.83	.75	.77	.66	.67	.49	.47	.42	.44
25	-.71	.52	.90	.91	.84	.83	.68	.68	.56	.57	.46	.48	.38	.40	.29	.29	.27	.27	.27	.27
26	1.00	.78	.98	.98	.97	.98	.94	.94	.89	.89	.83	.84	.77	.78	.68	.69	.51	.49	.43	.45
27	.71	.74	.97	.97	.97	.97	.92	.92	.85	.86	.78	.79	.70	.72	.59	.60	.43	.43	.40	.41
28	.09	.64	.95	.96	.94	.93	.85	.85	.75	.75	.65	.67	.55	.57	.41	.42	.33	.33	.33	.33
29	-.70	.52	.90	.91	.84	.83	.68	.68	.56	.57	.47	.48	.38	.40	.29	.29	.27	.27	.27	.27
30	-3.03	.26	.43	.46	.29	.27	.26	.26	.24	.24	.23	.23	.23	.22	.21	.21	.21	.21	.21	.21
31	-.02	.63	.95	.95	.93	.92	.83	.83	.73	.73	.62	.64	.52	.55	.39	.39	.32	.32	.32	.32
32	-1.60	.39	.76	.79	.59	.57	.45	.44	.38	.38	.32	.33	.29	.29	.24	.24	.23	.23	.24	.24
33	-.47	.56	.92	.93	.88	.87	.74	.74	.62	.63	.52	.54	.42	.44	.31	.32	.28	.28	.29	.29
34	.13	.65	.95	.96	.94	.94	.85	.85	.75	.76	.66	.67	.55	.58	.42	.43	.33	.34	.33	.34
35	-1.08	.46	.85	.87	.75	.74	.58	.58	.48	.48	.49	.41	.33	.34	.26	.26	.25	.25	.26	.25
36	-1.04	.47	.86	.88	.76	.75	.59	.59	.49	.49	.40	.41	.34	.35	.26	.26	.25	.25	.26	.26
37	-1.14	.45	.85	.86	.74	.72	.56	.56	.47	.47	.38	.40	.33	.33	.26	.26	.25	.25	.25	.25
38	.61	.72	.97	.97	.96	.96	.91	.91	.84	.85	.76	.78	.68	.70	.56	.57	.41	.41	.38	.39
39	-.24	.59	.93	.94	.91	.90	.79	.79	.68	.68	.57	.59	.47	.49	.35	.35	.30	.30	.30	.30
40	.47	.70	.96	.97	.96	.96	.90	.90	.82	.82	.73	.75	.64	.67	.51	.53	.38	.38	.37	.37
41	1.82	.88	.99	.99	.99	.99	.97	.97	.95	.95	.93	.93	.88	.89	.85	.86	.75	.73	.55	.60
42	-.89	.49	.88	.89	.80	.79	.63	.63	.52	.53	.43	.44	.36	.37	.27	.27	.26	.26	.26	.26
43	1.01	.78	.98	.98	.97	.98	.94	.94	.89	.90	.84	.84	.77	.78	.68	.70	.51	.49	.43	.45
44	-.11	.61	.94	.95	.92	.91	.81	.81	.70	.71	.60	.62	.50	.52	.37	.37	.31	.31	.31	.31
45	-.70	.52	.90	.91	.84	.83	.68	.68	.57	.57	.47	.48	.38	.40	.29	.29	.27	.27	.27	.27
46	-.28	.58	.93	.94	.90	.90	.78	.78	.67	.67	.56	.58	.46	.48	.34	.34	.29	.30	.30	.30
47	-.38	.57	.93	.93	.89	.88	.76	.76	.64	.65	.54	.56	.44	.46	.33	.33	.29	.29	.29	.29
48	1.29	.82	.98	.98	.98	.98	.95	.95	.92	.92	.87	.88	.81	.83	.75	.77	.60	.57	.47	.50
49	-2.34	.30	.58	.62	.38	.36	.32	.32	.28	.28	.25	.26	.25	.25	.22	.22	.22	.22	.22	.22
50	-.55	.54	.91	.92	.87	.86	.72	.72	.60	.61	.50	.52	.41	.43	.30	.31	.28	.28	.28	.28
51	.74	.74	.97	.97	.97	.97	.92	.92	.86	.86	.79	.80	.71	.73	.60	.61	.44	.43	.40	.41
52	.42	.69	.96	.97	.96	.95	.89	.89	.81	.81	.72	.74	.63	.65	.50	.51	.37	.38	.36	.37
53	-2.43	.30	.56	.60	.37	.35	.31	.31	.28	.28	.25	.25	.24	.24	.22	.22	.22	.22	.22	.22
54	-.91	.49	.88	.89	.80	.78	.63	.62	.52	.52	.42	.44	.35	.37	.27	.27	.26	.26	.26	.26
55	-1.23	.44	.83	.85	.71	.69	.54	.54	.45	.45	.37	.38	.32	.32	.25	.25	.25	.24	.25	.25
56	-1.73	.37	.73	.76	.55	.53	.42	.42	.36	.36	.30	.31	.28	.28	.23	.23	.23	.23	.24	.23
57	.27	.67	.96	.96	.95	.95	.87	.87	.78	.79	.69	.71	.59	.62	.46	.47	.35	.36	.35	.35
58	-.63	.53	.91	.92	.85	.84	.70	.70	.58	.59	.48	.50	.39	.41	.30	.30	.27	.27	.28	.28
59	-2.21	.32	.61	.65	.41	.39	.34	.34	.29	.30	.26	.27	.25	.25	.22	.22	.22	.22	.23	.22
60	-.57	.54	.91	.92	.86	.85	.71	.71	.60	.60	.49	.51	.40	.42	.30	.30	.28	.28	.28	.28
61	.22	.66	.96	.96	.94	.94	.87	.87	.77	.78	.68	.70	.58	.61	.44	.45	.35	.35	.34	.35
62	-.57	.54	.91	.92	.86	.85	.72	.71	.60	.60	.50	.51	.40	.42	.30	.30	.28	.28	.28	.28
63	-1.36	.42	.81	.83	.67	.65	.51	.50	.42	.42	.35	.36	.31	.31	.25	.25	.24	.24	.25	.24
64	-.23	.59	.93	.94	.91	.90	.79	.79	.68	.68	.57	.59	.47	.49	.35	.35	.30	.30	.30	.30
65	-1.68	.38	.74	.77	.57	.54	.43	.43	.36	.37	.31	.32	.28	.28	.23	.23	.23	.23	.24	.23
66	-.46	.56	.92	.93	.88	.87	.74	.74	.62	.63	.52	.54	.42	.44	.32	.32	.28	.28	.29	.29
67	-.08	.62	.94	.95	.92	.92	.82	.82	.71	.72	.61	.63	.50	.53	.37	.38	.31	.32	.32	.32
68	.54	.71	.97	.97	.96	.96	.90	.90	.83	.83	.75	.76	.66	.68	.53	.55	.39	.39	.38	.38
69	-.99	.48	.87	.88	.78	.76	.60	.60	.50	.50	.41	.42	.34	.35	.27	.27	.25	.25	.26	.26
70	-.06	.62	.94	.95	.92	.92	.82	.82	.72	.72	.61	.63	.51	.54	.38	.38	.31	.32	.32	.32
71	-.64	.53	.90	.92	.85	.84	.70	.69	.58	.59	.48	.50	.39	.41	.29	.30	.27	.27	.28	.28
72	.17	.65	.95	.96	.94	.94	.86	.86	.76	.77	.67	.68	.56	.59	.43	.44	.34	.34	.34	.34
73	.78	.75	.97	.97	.97	.97	.92	.92	.86	.87	.80	.81	.72	.74	.61	.63	.45	.44	.40	.42
74	.53	.71	.97	.97	.96	.96	.90	.90	.83	.83	.75	.76	.66	.68	.53	.55	.39	.39	.37	.38
75	-2.38	.30	.57	.61	.37	.35	.32	.31	.28	.28	.25	.25	.25	.24	.22	.22	.22	.22	.22	.22

(continued)

Table C *(continued)*
Ability Estimate ($\hat{\theta}$), Total Proportion Correct (T) and Proportion Correct for Each Student on Each of the Substrata (A, B) of the Nine-Stratum Test

			Stratum and Substratum																	
			1		2		3		4		5		6		7		8		9	
Student	$\hat{\theta}$	T	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
76	-.54	.54	.91	.92	.87	.86	.72	.72	.60	.61	.50	.52	.41	.43	.30	.31	.28	.28	.28	.28
77	-.50	.55	.92	.93	.87	.86	.73	.73	.61	.62	.51	.53	.42	.44	.31	.31	.28	.28	.29	.28
78	-.89	.49	.88	.89	.80	.78	.63	.63	.52	.53	.43	.44	.36	.37	.27	.27	.26	.26	.26	.26
79	-1.56	.39	.77	.80	.61	.58	.46	.46	.38	.39	.32	.33	.29	.29	.24	.24	.24	.23	.24	.24
80	-.97	.48	.87	.88	.78	.77	.61	.61	.50	.51	.41	.43	.35	.36	.27	.27	.26	.25	.26	.26
81	-3.08	.25	.42	.45	.28	.26	.26	.26	.24	.24	.23	.23	.23	.22	.21	.21	.21	.21	.21	.21
82	-.93	.49	.87	.89	.79	.78	.62	.62	.51	.52	.42	.44	.35	.36	.27	.27	.26	.26	.26	.26
83	.43	.70	.96	.97	.96	.96	.89	.89	.81	.82	.73	.74	.63	.66	.50	.52	.38	.38	.36	.37
84	.65	.73	.97	.97	.96	.97	.91	.91	.85	.85	.77	.78	.69	.71	.57	.59	.42	.41	.39	.40
85	-.60	.54	.91	.92	.86	.85	.71	.71	.59	.60	.49	.51	.40	.42	.30	.30	.27	.27	.28	.28
86	-.43	.56	.92	.93	.88	.88	.75	.75	.63	.64	.53	.54	.43	.45	.32	.32	.28	.29	.29	.29
87	-1.43	.41	.80	.82	.65	.63	.49	.49	.41	.41	.34	.35	.30	.30	.24	.24	.24	.24	.24	.24
88	-1.33	.43	.81	.83	.68	.66	.51	.51	.42	.43	.35	.36	.31	.31	.25	.25	.24	.24	.25	.24
89	-.87	.49	.88	.89	.81	.79	.64	.63	.53	.53	.43	.45	.36	.37	.27	.28	.26	.26	.26	.26
90	.39	.69	.96	.97	.95	.95	.89	.89	.81	.81	.72	.73	.62	.65	.49	.50	.37	.37	.36	.37
91	-.35	.57	.93	.94	.89	.89	.77	.76	.65	.66	.54	.56	.44	.47	.33	.33	.29	.29	.29	.29
92	-.03	.62	.95	.95	.93	.92	.83	.83	.72	.73	.62	.64	.51	.54	.38	.39	.32	.32	.32	.32
93	-.56	.69	.96	.97	.95	.95	.89	.89	.81	.81	.72	.73	.62	.65	.49	.50	.37	.37	.36	.37
94	-2.40	.30	.56	.60	.37	.35	.32	.31	.28	.28	.25	.25	.24	.24	.22	.22	.22	.22	.22	.22
95	-.88	.49	.88	.89	.80	.79	.64	.63	.52	.53	.43	.45	.36	.37	.27	.28	.26	.26	.26	.26
96	-.30	.58	.93	.94	.90	.89	.78	.78	.66	.67	.56	.57	.45	.48	.34	.34	.29	.30	.30	.30
97	-1.84	.36	.70	.74	.51	.49	.40	.40	.34	.34	.29	.30	.27	.27	.23	.23	.23	.23	.23	.23
98	-.64	.53	.90	.92	.85	.84	.70	.70	.58	.59	.48	.50	.39	.41	.30	.30	.27	.27	.28	.28
99	-1.26	.44	.83	.85	.70	.68	.53	.53	.44	.45	.36	.38	.32	.32	.25	.25	.24	.24	.25	.25
100	-.51	.55	.92	.93	.87	.86	.73	.73	.61	.62	.51	.53	.41	.43	.31	.31	.28	.28	.28	.28
101	.96	.77	.98	.98	.97	.98	.94	.94	.89	.89	.83	.84	.76	.77	.66	.68	.50	.48	.43	.45
102	-2.10	.33	.64	.68	.44	.42	.36	.35	.31	.31	.27	.27	.26	.26	.22	.22	.22	.22	.23	.22
103	1.36	.83	.98	.98	.98	.99	.96	.96	.93	.93	.88	.89	.83	.84	.77	.78	.63	.60	.48	.51
104	-.80	.51	.89	.90	.82	.81	.66	.65	.54	.55	.45	.46	.37	.38	.28	.28	.26	.26	.27	.27
105	-.75	.51	.89	.91	.83	.82	.67	.67	.55	.56	.46	.47	.38	.39	.29	.29	.27	.27	.27	.27
106	-.79	.51	.89	.90	.82	.81	.66	.66	.54	.55	.45	.46	.37	.38	.28	.28	.26	.26	.27	.27
107	-.23	.59	.93	.94	.91	.90	.79	.79	.68	.68	.57	.59	.47	.49	.35	.35	.30	.30	.30	.30
108	.27	.67	.96	.96	.95	.95	.87	.87	.78	.79	.69	.71	.59	.62	.46	.47	.35	.35	.35	.35
109	-.39	.57	.92	.93	.89	.88	.76	.76	.64	.65	.54	.55	.44	.46	.32	.33	.29	.29	.29	.29
110	-.38	.57	.93	.93	.89	.88	.76	.76	.64	.65	.54	.56	.44	.46	.33	.33	.29	.29	.29	.29
111	-.68	.52	.90	.91	.84	.83	.69	.68	.57	.58	.47	.49	.39	.40	.29	.29	.27	.27	.27	.27
112	-.27	.59	.93	.94	.90	.90	.78	.78	.67	.68	.56	.58	.46	.49	.34	.35	.30	.30	.30	.30
113	-.85	.50	.88	.90	.81	.79	.64	.64	.53	.54	.43	.45	.36	.37	.28	.28	.26	.26	.27	.26
114	1.06	.79	.98	.98	.98	.98	.94	.94	.90	.90	.84	.85	.78	.79	.69	.71	.53	.51	.44	.46
115	-.82	.50	.89	.90	.82	.80	.65	.65	.54	.54	.44	.46	.37	.38	.28	.28	.26	.26	.27	.27
116	.42	.69	.96	.97	.96	.95	.89	.89	.81	.81	.72	.74	.63	.65	.50	.51	.37	.38	.36	.37
117	.89	.76	.97	.98	.97	.97	.93	.93	.88	.88	.82	.82	.74	.76	.64	.66	.48	.46	.42	.43
118	-3.40	.24	.37	.39	.26	.24	.24	.24	.23	.23	.22	.22	.22	.22	.21	.21	.21	.21	.21	.21
119	-.88	.49	.88	.89	.80	.79	.64	.63	.52	.53	.43	.45	.36	.37	.27	.28	.26	.26	.26	.26
120	-.49	.55	.92	.93	.88	.87	.74	.73	.62	.62	.51	.53	.42	.44	.31	.31	.28	.28	.29	.29
121	.31	.71	.97	.97	.96	.96	.90	.90	.82	.83	.74	.76	.65	.68	.52	.54	.39	.39	.37	.38
122	2.07	.90	.99	.99	.99	.99	.98	.98	.96	.96	.94	.95	.90	.91	.89	.89	.81	.79	.60	.64
123	-1.95	.34	.68	.71	.48	.46	.38	.38	.32	.33	.28	.29	.27	.27	.23	.23	.23	.22	.23	.23
124	.75	.74	.97	.97	.97	.97	.92	.92	.86	.86	.79	.80	.71	.73	.60	.62	.44	.43	.40	.41
125	-2.08	.33	.65	.68	.45	.42	.36	.36	.31	.31	.27	.28	.26	.26	.22	.22	.22	.22	.23	.23
126	-.98	.48	.87	.88	.78	.76	.61	.60	.50	.51	.41	.43	.34	.35	.27	.27	.25	.25	.26	.26
127	.82	.75	.97	.97	.97	.97	.93	.93	.87	.87	.80	.81	.73	.75	.62	.64	.46	.45	.41	.42
128	-.14	.61	.94	.95	.92	.91	.81	.81	.70	.70	.59	.61	.49	.52	.36	.37	.31	.31	.31	.31
129	-1.71	.37	.74	.76	.56	.53	.42	.42	.36	.36	.30	.31	.28	.28	.23	.23	.23	.23	.24	.23
130	-.16	.60	.94	.95	.92	.91	.81	.81	.69	.70	.59	.61	.49	.51	.36	.37	.31	.31	.31	.31
131	.06	.64	.95	.96	.93	.93	.84	.84	.74	.75	.64	.66	.54	.57	.40	.41	.33	.33	.33	.33
132	-.57	.54	.91	.92	.86	.85	.72	.71	.60	.60	.49	.51	.40	.42	.30	.30	.28	.28	.28	.28
133	-1.32	.43	.81	.84	.68	.66	.51	.51	.43	.43	.35	.37	.31	.31	.25	.25	.24	.24	.25	.24
134	-2.08	.33	.65	.68	.45	.42	.36	.36	.31	.31	.27	.28	.26	.26	.22	.22	.22	.22	.23	.23
135	-1.01	.47	.86	.88	.77	.75	.60	.59	.49	.50	.40	.42	.34	.35	.27	.27	.25	.25	.26	.26
136	1.87	.88	.99	.99	.99	.99	.97	.97	.96	.96	.93	.93	.89	.89	.86	.87	.77	.74	.56	.61
137	-.73	.52	.90	.91	.84	.82	.68	.67	.56	.57	.46	.48	.38	.39	.29	.29	.27	.27	.27	.27
138	.35	.68	.96	.96	.95	.95	.88	.88	.80	.80	.71	.72	.61	.64	.48	.49	.36	.37	.36	.36
139	.25	.67	.96	.96	.95	.94	.87	.87	.78	.78	.69	.70	.59	.61	.45	.46	.35	.35	.34	.35
140	-1.05	.47	.86	.87	.76	.74	.59	.58	.48	.49	.40	.41	.34	.35	.26	.26	.25	.25	.26	.25
141	-.74	.51	.90	.91	.83	.82	.67	.67	.56	.56	.46	.48	.38	.39	.29	.29	.27	.27	.27	.27
142	.75	.74	.97	.97	.97	.97	.92	.92	.86	.86	.79	.80	.71	.73	.60	.62	.44	.43	.40	.41
143	-.47	.56	.92	.93	.88	.87	.74	.74	.62	.63	.52	.53	.42	.44	.31	.32	.28	.28	.29	.29
144	-.96	.48	.87	.88	.79	.77	.61	.61	.51	.51	.41	.43	.35	.36	.27	.27	.26	.25	.26	.26
145	-1.00	.47	.86	.88	.77	.76	.60	.60	.49	.50	.41	.42	.34	.35	.27	.27	.25	.25	.26	.26
146	-1.01	.47	.86	.88	.77	.75	.60	.59	.49	.50	.40	.42	.34	.35	.27	.27	.25	.25	.26	.26
147	-.75	.51	.89	.91	.83	.82	.67	.67	.55	.56	.46	.47	.38	.39	.28	.29	.27	.27	.27	.27
148	-.14	.61	.94	.95	.92	.91	.81	.81	.70	.70	.59	.61	.49	.52	.36	.37	.31	.31	.31	.31
149	1.29	.82	.98	.98	.98	.98	.95	.95	.92	.92	.87	.88	.82	.83	.75	.77	.60	.57	.47	.50
150	-.04	.62	.94	.95	.93	.92	.83	.83	.72	.73	.62	.64	.51	.54	.38	.39	.32	.32	.32	.32
151	-.13	.61	.94	.95	.92	.91	.81	.81	.70	.71	.60	.62	.49	.52	.37	.37	.31	.31	.31	.31

Navy

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

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

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

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 CDR Robert S. Kennedy
Naval Aerospace Medical and Research Lab
Box 29407
New Orleans, LA 70189

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

1 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 CAPT Richard L. Martin
USS Francis Marion (LPA-249)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

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

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 OFFICE OF CIVILIAN PERSONNEL
(CODE 26)
DEPT. OF THE NAVY
WASHINGTON, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION & TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Code 436
Office of Naval Research
Arlington, VA 22217

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs (Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc (OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

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

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

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

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

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

Army

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH DUSEK
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. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. 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 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

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

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

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

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

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

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

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

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

Other DoD

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM
Ft. 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 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

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

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

Non Govt

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

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

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

1 Dr. Warner Birice
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

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

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

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

1 Dr. William Coffman
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Hans Cronbag
Education Research Center
University of Leyden
Boerhaavelaan 2
Leyden
The NETHERLANDS

1 MAJOR I. M. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

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 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Gerhardt Fischer
Liebigasse 5
Vienna 1010
Austria

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

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ross Greene
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

1 Dr. Alan Gross
Center for Advanced Study in Education
City University of New York
New York, NY 10036

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

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

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

1 Dr. Carl J. Jensema
Gallaudet College
Kendall Green
Washington, DC 20002

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. John A. Keats
University of Newcastle
Newcastle, New South Wales
AUSTRALIA

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A OK2

1 Dr. Michael Levine
Department of Educational Psychology
University of Illinois
Champaign, IL 61820

1 Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

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

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77025

1 Dr. Sam Mayo
Loyola University of Chicago
Chicago, IL 60601

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

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

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Donald Rubin
Educational Testing Service
Princeton, NJ 08450

1 Dr. Larry Rudner
Gallaudet College
Kendall Green
Washington, DC 20002

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 Dr. Kazao Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 982
JAPAN

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

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

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

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

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

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

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

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

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

1 Dr. Robert Tsutakawa
Dept. of Statistics
University of MIssouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, West Germany D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.
79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455