

AN ADAPTIVE TESTING STRATEGY FOR ACHIEVEMENT TEST BATTERIES

Joel M. Brown
and
David J. Weiss

RESEARCH REPORT 77-6
OCTOBER 1977

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

Prepared under contract No. N00014-76-C-0627, NR150-389
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

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

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM												
1. REPORT NUMBER Research Report 77-6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER												
4. TITLE (and Subtitle) An Adaptive Testing Strategy for Achievement Test Batteries		5. TYPE OF REPORT & PERIOD COVERED Technical Report												
		6. PERFORMING ORG. REPORT NUMBER												
7. AUTHOR(s) Joel M. Brown 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: RR-42-04 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 October 1977												
		13. NUMBER OF PAGES 48												
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) <table border="0"> <tr> <td>testing</td> <td>sequential testing</td> <td>programmed testing</td> </tr> <tr> <td>achievement testing</td> <td>branched testing</td> <td>response-contingent testing</td> </tr> <tr> <td>computerized testing</td> <td>individualized testing</td> <td>automated testing</td> </tr> <tr> <td>adaptive testing</td> <td>tailored testing</td> <td></td> </tr> </table>			testing	sequential testing	programmed testing	achievement testing	branched testing	response-contingent testing	computerized testing	individualized testing	automated testing	adaptive testing	tailored testing	
testing	sequential testing	programmed testing												
achievement testing	branched testing	response-contingent testing												
computerized testing	individualized testing	automated testing												
adaptive testing	tailored testing													
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An adaptive testing strategy is described for use with achievement tests which cover multiple content areas. The testing strategy combines adaptive item selection both within and between the subtests in the multiple-subtest battery. A real-data simulation was conducted in order to compare the results from adaptive testing with those from conventional testing, in terms of test information and test length. Data for the simulation consisted of test results for 365 fire-control technicians on a paper-and-pencil administration of a 232-item achievement test which was divided into 12 subtests, each														

covering a different content area. Correlations between subtest scores from adaptive and conventional testing were .90 or higher for eleven of the twelve content areas. An information analysis showed that for all 12 subtests, the subtest information curves from adaptive testing were essentially identical to the corresponding subtest information curves from conventional testing. On the average, the number of items administered with adaptive testing was half as many as was required with conventional testing; the shortest adaptive test battery used 18% of the total number of items in the conventional test, while the longest used 80%. The adaptive testing strategy, therefore, provided a considerable reduction in test length and virtually no loss in precision of measurement when compared with the conventional administration of the achievement test battery.

CONTENTS

Introduction	1
Method	2
Purpose	2
Procedure	2
Test Items and Subjects	2
Item Parameterization	2
Adaptive Testing Strategy	3
Intra-Subtest Branching	3
Item selection	3
Estimation of θ	4
Termination criteria	5
Inter-Subtest Branching	6
Subtest ordering	6
Differential entry points	7
Conventional Test	9
Data Analysis	9
Correlation Analysis	9
Information Analysis	9
Results	10
Preliminary Results	10
Item parameterization	10
Subtest ordering	11
Comparison of Adaptive and Conventional Tests	12
Test length	12
Correlation Analysis	13
Information Analysis	14
Termination criterion .001	14
Termination criterion .01	15
Discussion	18
Applicability of the ICC Model	19
Intra-Subtest Branching	20
Inter-Subtest Branching	21
Conclusions	22
References	24
Appendix A: Illustration of Intra-Subtest Adaptive Branching	26
Appendix B: Supplementary Tables	29
Appendix C: Supplementary Figures	44

AN ADAPTIVE TESTING STRATEGY FOR ACHIEVEMENT TEST BATTERIES

Modern test theory (latent trait theory) has provided the framework for a growing body of research in ability measurement through adaptive testing. Weiss and Betz (1973) have presented a comprehensive review of adaptive testing which suggested that adaptive testing can considerably reduce testing time, while concurrently yielding scores of higher reliability and validity than those yielded by conventional tests. During the past several years, a number of studies have been published which were concerned with applications of different adaptive testing strategies in the ability domain (e.g., Betz & Weiss, 1974, 1975; Larkin & Weiss, 1974, 1975; Lord, 1977; McBride & Weiss, 1976; Urry, 1977; Vale & Weiss, 1975). Each of these studies, as well as all the previous research in adaptive testing (Weiss & Betz, 1973), has been concerned with tests which covered only a single content area. Thus, all of the branching procedures implemented for the adaptive selection of items to be administered to a testee have been designed exclusively for intra-test branching. That is, items were selected within a single, presumably unidimensional, content area.

Recent studies (e.g., Bejar, Weiss, & Gialluca, 1977; Bejar, Weiss, & Kingsbury, 1977) have demonstrated that unidimensional approaches to intra-test adaptive testing are useful for measurement in the achievement domain. Frequently, however, achievement tests span several content areas. Consequently, in many cases the assumption of a single dimension may not be appropriate. For these kinds of achievement tests, or for achievement test batteries covering a number of separable content areas for which separate scores are required, none of the existing adaptive strategies described by Weiss (1974) are directly applicable.

There are two reasons why many of the adaptive testing strategies developed for single-content area ability tests may not be appropriate for achievement tests which cover several content areas. The first reason is that although the unidimensional branching models can be applied to separate content areas, they are not designed to take into account the information available between content areas. The second, and more practical, reason is that it might not be possible to generate relatively large numbers of items, such as those required for many adaptive testing strategies, within one content area in an achievement test. Urry (1977) has suggested that item pools to be used in adaptive testing with Owen's (1975) Bayesian testing strategy should include a minimum of 100 items to measure one dimension. Although there are no firm guidelines for other adaptive testing strategies, it is evident that they will function best with large item pools. Thus, application of these strategies to an achievement test battery of five subtests would require the test constructor to assemble 500 items with good psychometric qualities. Frequently, this is not possible. Consequently, in the application of adaptive testing to the unique problems in the measurement of achievement, an important research issue is the identification of adaptive testing strategies which make efficient use of existing item pools, rather than requiring the re-design of test item pools to meet the requirements of specific adaptive testing strategies.

The present paper describes an adaptive testing strategy which can be used in achievement tests with relatively small numbers of items. The strategy is designed for achievement test batteries or achievement tests with multiple content areas. It incorporates both intra-subtest branching and inter-subtest branching in order to efficiently adapt the test battery to each individual testee. The adaptive testing strategy is applied to a test battery and evaluated in terms of:

1. The reduction in number of items administered,
2. Correlations of ability estimates with those derived from conventional administration of the test battery, and
3. The effects of adaptive administration on the psychometric information in the test scores.

METHOD

Purpose

The purpose of this study was to develop and evaluate an efficient and generalizable adaptive testing strategy for an achievement test battery comprised of a number of subtests. The adaptive testing strategy developed is designed to operate within a fixed item pool containing a relatively small number of items for each subtest. Real data simulation techniques (Weiss & Betz, 1973, pp. 11-12) were used. That is, the adaptive testing strategy was applied to item response data obtained from the administration of an achievement test battery which had been previously administered conventionally by paper-and-pencil. Results for the conventional testing strategy were compared with those for the adaptive testing strategy in terms of both test information and test length.

Procedure

Test Items and Subjects

Achievement test data were provided by the Personnel and Training Evaluation Program (PTEP) of the Naval Guided Missile School at Dam Neck, Virginia.¹ These data were from a systems achievement test (SAT F17603) battery administered to 365 fire control technicians. The test battery included twelve subtests, each covering knowledge areas for different equipment or subject matter. Table 1 shows the content and number of items in each subtest. The test battery was administered in one booklet containing 232 items. The number of items per subtest ranged from 10 to 32; all of the items were multiple-choice with four response choices. The data provided by PTEP consisted of an identification number for each testee, the testee's number correct score on each of the twelve subtests, and correct-incorrect item responses for each of the 232 items.

Item Parameterization

Items were parameterized using Urry's ESTEM computer program (see Urry, 1976, p. 99) for latent trait item parameterization employing the three-parameter normal ogive model. This program provided estimates of the item discrimination (a), item difficulty (b), and guessing (c) parameters. The items for

¹Data were generously supplied by Lieutenant Commander Lee J. Walker of PTEP.

Table 1
Number of Items in Each Subtest

Subtest	Content	No. of Items
A	Fire control system casualty procedures	10
B	Optical alignment group	10
C	Control console and power subsystem	18
D	Platform positioning equipment	22
E	Multiplexed equipment	18
F	Digital control computer and software	18
G	Digital control computer--operator interface	14
H	Magnetic disk file	12
I	Digital control computer--missile interface	24
J	Guidance and guidance testing	29
K	MTRE MKG MOD3	32
L	Spare guidance temperature monitor	25
Total		232

each subtest were parameterized independently of items in other subtests.

Urry's item parameterization program calculates item parameter estimates using a two-phase procedure. In the first phase, initial item parameter estimates are determined for all items. However, item parameters 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 phase, item parameters are recomputed for all items which are not excluded by the criteria applied in the first phase. In this phase, item parameter estimates are reported without restrictions (e.g., c may be greater than .30 for some items in the second phase) for all items not excluded in the first phase.

Adaptive Testing Strategy

The adaptive testing procedure was developed in order to reduce to a minimum the number of items administered to each individual with as little impact as possible upon the measurement characteristics of the test battery. Both intra-subtest adaptive branching and inter-subtest adaptive branching were used in the development of the procedure.

Intra-Subtest Branching

Item selection. The basic concept for intra-subtest adaptive branching was that the order in which the items were to be administered was to be dependent upon values of the item information curve as defined by Birnbaum (1968). For

each item in each subtest, item information values were computed using Equation 1 (Birnbaum, 1968, p. 462):

$$I_g(\theta) = (1-c_g)D^2\alpha_g^2\psi^2[DL_g(\theta)]/\{\psi DL_g(\theta)\} + c_g\psi^2[-DL_g(\theta)] \quad [1]$$

where

$D = 1.7$; this is the scaling factor which maximizes agreement between the normal ogive and logistic latent trait models;

$$L_g(\theta) = \alpha_g(\theta - b_g);$$

ψ = the logistic probability density function;

Ψ = the cumulative logistic distribution function;

and values for the parameters α_g , b_g , and c_g , were derived for each of the g items in a subtest from the results of the item parameterization phase.

The information values for each item, $I_g(\theta)$, were computed for values of θ ranging from -3.0 to +3.0 in steps of .2 for each item in each subtest.

Items were selected within a subtest for each testee by computing the value of all item information curves at the current estimated achievement level ($\hat{\theta}$) for that testee using Equation 1. The item selected for administration was the item which had the highest information value at the testee's current level of $\hat{\theta}$. Once an item was administered to a testee, it was eliminated from the subtest pool of available items for that testee.

Estimation of θ . Owen's (1975) Bayesian scoring procedure was used for this simulation study. This scoring procedure provides an achievement level estimate ($\hat{\theta}$) after each m th test item is administered. The procedure begins with a prior estimate of $\hat{\theta}_m$ and its variance (σ_m^2). For the first item of the first subtest administered ($m=1$), these were 0.00 and 1.00, respectively. An item is administered and scored as correct (1) or incorrect (0). For a correct response, the revised estimate of $\hat{\theta}$ is determined by Equation 2,

$$\hat{\theta}_{m+1} = E(\theta|1) = \hat{\theta}_m + (1-c_g) \left(\sqrt{\frac{\sigma_m^2}{\alpha_g^2 + \sigma_m^2}} \right) \left(\frac{\phi(D)}{c_g + (1-c_g)\Phi(-D)} \right) \quad [2]$$

and its variance by Equation 3,

$$\sigma_{m+1}^2 = \text{var}(\theta|1) = \sigma_m^2 \left\{ 1 - \left(\frac{1-c_g}{1 + \frac{1}{\alpha_g^2 \sigma_m^2}} \right) \left(\frac{\phi(D)}{A} \right) \left(\frac{(1-c_g)\phi(D)}{A} - D \right) \right\}. \quad [3]$$

For an incorrect response, the revised estimate of $\hat{\theta}$ is determined by Equation 4,

$$\hat{\theta}_{m+1} = E(\theta|0) = \hat{\theta}_m - \left(\sqrt{\frac{\sigma_m^2}{\frac{1}{\alpha_g^2} + \sigma_m^2}} \right) \left(\frac{\phi(D)}{\Phi(D)} \right) \quad [4]$$

and its variance by Equation 5,

$$\sigma_{m+1}^2 = \text{var}(\theta|0) = \sigma_m^2 \left\{ 1 - \left(\frac{\phi(D)}{1 + \frac{1}{\alpha_g^2 \sigma_m^2}} \right) \left(\frac{\frac{\phi(D)}{\Phi(D)} + D}{\Phi(D)} \right) \right\}. \quad [5]$$

In Equations 2 through 5 (adapted from Owen, 1975, p. 353)

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

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

$$D = \sqrt{\frac{b_g - \hat{\theta}_m}{\frac{1}{\alpha_g^2} + \sigma_m^2}}; \quad [6]$$

$$A = c_g + (1 - c_g)\Phi(-D); \text{ and} \quad [7]$$

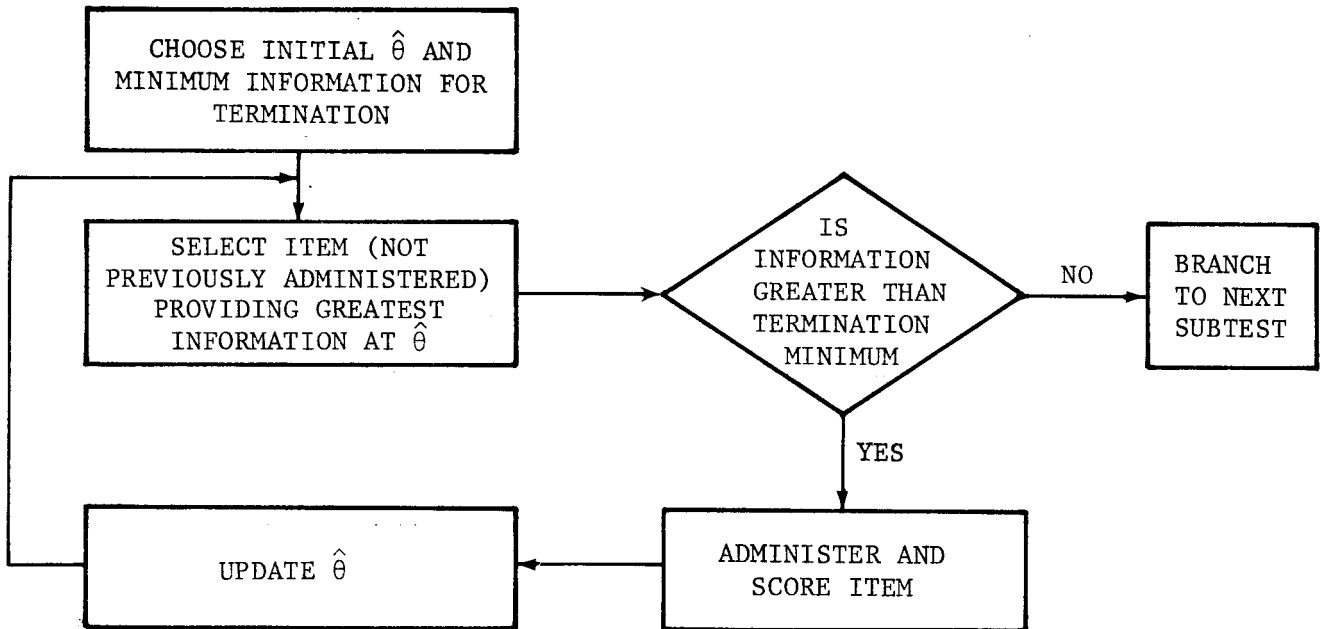
α_g , b_g and c_g are the item parameter estimates.

The updated estimates of $\hat{\theta}$ from either Equations 2 or 4, along with their associated variances, are used as the prior estimates of $\hat{\theta}$ for the selection of the next test item, which is based on the maximum information rule described above. The next item is administered; and a new value of $\hat{\theta}$ is determined, which is then used to select the next item. This procedure is repeated until a termination criterion is reached.

Termination criteria. Two criteria were used in determining when administration of items within a subtest should be stopped: 1) when all of the remaining items provided less than a pre-determined small amount of information; or 2) when the within-subtest item pool was exhausted. Testing was terminated for a given testee at the first occurrence of one of these criteria within a given subtest. In applying the first criterion, two arbitrarily small values of information were studied; testing was terminated when there was no item available which provided an information value greater than .01 or .001 at a given testee's current level of $\hat{\theta}$.

Figure 1 diagrammatically summarizes the intra-subtest branching procedure. Appendix A gives an illustration of this procedure, using six items from Subtest 1.

Figure 1
Intra-subtest Branching Scheme



Inter-Subtest Branching

Subtest ordering. The order of administration for the various subtests was chosen to take maximum advantage of the intercorrelations among them, thereby utilizing the redundant information in previously administered subtests. This was accomplished through linear multiple regression. First, the number correct subtest scores for the twelve subtests were intercorrelated, and the highest bivariate correlation was chosen from the intercorrelation matrix. One of these two subtests was arbitrarily designated to be administered first; the other was designated to be administered second.

Multiple correlations were then computed using the subtests previously designated first and second as predictor variables. Each of the ten remaining subtests, in turn, was designated as the criterion variable. Of these ten subtests, the one which had the highest multiple correlation with the first and second subtests was designated as the third subtest. This procedure was repeated to select the fourth subtest for the adaptive administration, computing multiple

correlations with the first three subtests as predictor variables and each of the remaining nine subtests, in turn, as the criterion variable. That subtest having the highest multiple correlation with the first three subtests was selected as the fourth subtest to be administered. By adding one subtest to the predictor set at each subsequent stage, this procedure was continued until all twelve subtests were ordered.

As a result of this procedure, the order in which the subtests were administered was the same for all testees. However, the selection of items within each subtest and the order in which those items were administered varied with testees as a function of the amount of item information provided at the testee's current achievement estimate.

Differential subtest entry points. An important feature of the adaptive testing strategy implemented in this study was that after the first subtest, each testee's entry points for the second and subsequent subtests were differentially determined. For the first subtest, each testee's achievement level was assumed to be $\hat{\theta}=0.00$. That is, having no previous information on which to base an estimate of the testee's achievement level, the initial item chosen from the first subtest for administration was the item which provided the most information for an estimated achievement level at the mean of the $\hat{\theta}$ distribution. Thus, all testees began the first subtest with the same test item.

The entry point into the item pool for the second subtest was determined from both the examinee's $\hat{\theta}$ at the end of the first subtest and the bivariate regression of scores from Subtest 1 on Subtest 2. This regression equation was based not only on scores for the items administered adaptively, but also on the correlations derived from number correct scores for all items in each of the subtests. The first item to be administered for a testee in the second subtest was determined from information provided by evaluating Equation 8.

$$\hat{\theta}_{2E} = B_{12}\hat{\theta}_1 + A \quad [8]$$

where

$\hat{\theta}_{2E}$ is the first $\hat{\theta}$ used for selection of the first test item in Subtest 2,

$\hat{\theta}_1$ is the final $\hat{\theta}$ for a testee at completion of the adaptive administration of items in Subtest 1,

B_{12} is the bivariate regression coefficient for the regression of Subtest 2 on Subtest 1, and

A is the regression constant.

The entry achievement level estimate, $\hat{\theta}_{2E}$, computed as $\hat{\theta}_m$ by Equations 2 and 4, was used for selecting the first item to be administered in Subtest 2. The variance of this estimate (σ_m^2 in Equations 3 and 5) was determined by Equation 9, which is the formula for the squared standard error of estimate

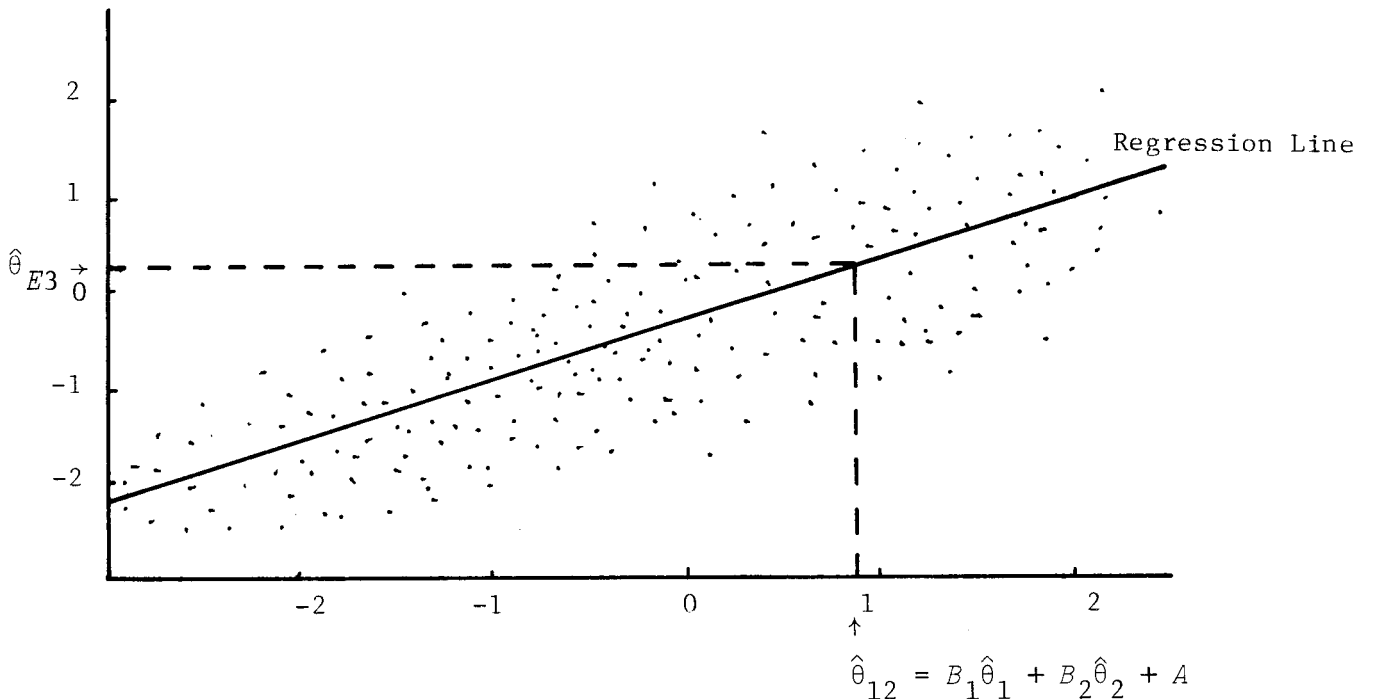
in bivariate regression (adapted from Glass & Stanley, 1970, p. 143):

$$s_e^2 = s_2^2 + r_{12}^2 s_2^2 - 2B_1 s_{12} = s_2^2 (1 - r_{12}^2), \quad [9]$$

where the subscripts 1 and 2 represent the first and second subtests.

Determination of the entry point for the third and subsequent subtests was merely a generalization of the method used for the second subtest. The testee's achievement level estimates from Subtest 1 ($\hat{\theta}_1$) and Subtest 2 ($\hat{\theta}_2$) were entered into the multiple regression equation for predicting Subtest 3 scores from scores on Subtests 1 and 2. This generated an estimated subtest score for an individual ($\hat{\theta}_{E3}$), which was used as the initial prior achievement level estimate ($\hat{\theta}_m$) for intra-subtest branching in Subtest 3. The squared standard error of estimate from the multiple regression of Subtests 1 and 2 on Subtest 3 was used as the initial prior variance (σ_m^2) of the Bayesian achievement level estimate for Subtest 3. Figure 2 illustrates this differential entry point procedure.

Figure 2
Estimation of Initial Achievement Level Estimate for Subtest 3 ($\hat{\theta}_{E3}$)
From the Multiple Regression of Subtest 1 ($\hat{\theta}_1$) and Subtest 2 ($\hat{\theta}_2$)



The inter-subtest branching regression procedure was used for entry into each of the remaining subtests. Each subsequent regression equation was based on the achievement estimates from each of the previously administered subtests. A testee's achievement level estimates for each subtest, based on the multiple regression of all previous subtests on a new subtest, was used as the initial Bayesian prior $\hat{\theta}$ for intra-subtest branching within that subtest. Item selection and scoring within subsequent subtests was then based on the intra-subtest branching procedures described earlier.

Conventional Test

A conventional test was used for comparison with the adaptive testing strategy. The subtests were administered in the same order for both the conventional and adaptive strategies. In the conventional strategy, all items within each subtest were administered sequentially, so that all testees took the same items in the same order. Hence, there was no differential entry for the conventional strategy. In addition, all testees completed all items, which is typical in conventional testing.

In order to facilitate comparison of results with the adaptive strategy, Bayesian scoring was employed for the conventional test. A mean of 0.0 and a variance of 1.0 were used as the initial prior achievement estimate of the Bayesian score for each subtest.

Data Analysis

The basic question examined in this study was whether the number of items administered could be reduced through adaptive testing without significantly changing the characteristics of the test scores. The effects of reducing the number of items by the adaptive testing item selection procedure were evaluated by means of both a correlational analysis and an information analysis.

Correlation Analysis

Early research comparing single test adaptive testing strategies with conventional testing strategies (See Betz & Weiss, 1973, 1974; Larkin & Weiss, 1974, 1975; Vale & Weiss, 1975; Weiss, 1973) demonstrated that adaptive tests resulted in test scores highly correlated with conventional test scores, even though the adaptive tests required substantially fewer items. Consequently, in the present study Pearson product-moment correlations were computed between subtest achievement level estimates ($\hat{\theta}$) from the conventional and adaptive testing procedures in order to examine the extent of the relationship between the scores. These were computed separately for each of the twelve subtests. High correlations between the scores would suggest that the tests ranked the examinees in a similar order along the achievement continuum.

Information Analysis

Information analyses were conducted in order to compare the adaptive and conventional testing strategies as a function of achievement levels. Test information values for different testing strategies at different levels on the

achievement continuum provide an indication of their relative degree of precision of measurement (Birnbaum, 1968).

Estimated test information curves were generated separately for each subtest for both conventional and adaptive testing strategies. In the conventional testing strategy, an examinee's subtest information value was computed by summing the item information values at the examinee's final estimated achievement level ($\hat{\theta}$) for that subtest. An estimated information curve was plotted for the total group of examinees from their individual achievement level estimates and corresponding information values. For a conventional test this is equivalent to computing the test information function using the item parameters a , b , and c , as suggested by Birnbaum (1968, pp. 454-464).

Estimated subtest information curves were generated similarly for the adaptive testing strategy. The estimated value of test information was computed at each testee's final achievement estimate for the subtest by summing the information values at that $\hat{\theta}$ for the particular subset of items administered to that testee. Thus, for both adaptive and conventional testing, each test information value was computed at the final value of $\hat{\theta}$ for the subtest, based on the information provided by the items actually administered.

RESULTS

Preliminary Results

Item parameterization. Table 2 presents means and standard deviations for estimates of the latent trait item discrimination (a), difficulty (b), and guessing (c) parameters for the items in the twelve subtests. Complete distributions of individual item parameter estimates by subtest are shown in Appendix Table B-1.

Table 2
Means and Standard Deviations of Normal Ogive Item Discrimination (a),
Difficulty (b), and Guessing (c) Parameters for 12 Subtests

Subtest	Number of Items		a		b		c	
	Avail- able	Parame- terized	Mean	S.D.	Mean	S.D.	Mean	S.D.
A	10	10	1.90	.62	.06	1.03	.52	.11
B	10	10	2.12	.86	.31	1.29	.53	.18
C	18	15	1.80	.56	.54	1.30	.55	.08
D	22	19	1.60	.60	.43	1.28	.47	.08
E	18	17	1.57	.65	.74	1.32	.47	.10
F	18	18	1.58	.43	1.19	1.45	.56	.09
G	14	13	1.98	.94	1.20	1.26	.52	.18
H	12	12	2.12	.90	.84	1.10	.43	.10
I	24	22	1.49	.59	.88	1.36	.43	.10
J	29	23	1.66	.57	1.28	1.12	.44	.14
K	32	24	1.48	.61	.91	1.39	.43	.14
L	25	18	1.73	.58	1.44	1.34	.52	.17

From the total item pool of 232 items, item parameter estimates were obtained for 201 items (87%). Several of the subtests (A, B, F, H) did not lose any items in the calibration process; the largest loss (28% of the original number of items) occurred for Subtest L.

Mean item discrimination (α) ranged from 1.48 for Subtest K to 2.12 for Subtest H, while mean item difficulty ranged from .06 for Subtest A to 1.44 for Subtest L. Mean estimates for the c parameters of these four-alternative multiple choice items were relatively high, ranging from a low of .43 to a high of .56.

Subtest ordering. Table 3 shows the product-moment intercorrelations among subtest scores for the twelve content area subtests used to determine the order in which the subtests would be administered in the adaptive test. The highest bivariate correlation (.53) was between content areas C and K, which were designated Subtest 1 and Subtest 2, respectively.

Table 3
Intercorrelations Among Content Area Scores

	A	B	C	D	E	F	G	H	I	J	K
B	31										
C	40	37									
D	36	40	46								
E	37	37	48	38							
F	30	26	36	39	38						
G	30	38	41	36	46	35					
H	25	29	29	28	35	30	36				
I	23	33	42	48	47	45	41	28			
J	19	35	27	33	28	33	33	27	40		
K	42	33	53	39	41	30	37	28	35	27	
L	27	27	22	14	29	16	27	26	26	31	26

Note. Decimal points omitted.

Table 4 contains the multiple correlations for each subtest predicted from all previous subtests and shows the ordering of subtests based on the multiple correlations. The second column of Table 4 shows the order sequence numbers for the tests, based on their ordering by the multiple correlation procedure. These order sequence numbers are used throughout the remainder of this report to identify the subtests. The multiple correlations reported in Table 4 ranged from a low of .22, for predicting the score on Subtest L (12) from the score on Subtest C (1), to a high of .57, for predicting performance on Subtest D (5) from performance on the best weighted linear combination of Subtests C, K, E, I (1,2,3,4).

The inter-subtest multiple correlations shown in Table 4 were not high enough to justify applying a unidimensional adaptive testing strategy model across subtests; instead, a multi-subtest branching strategy was developed and implemented as a more appropriate procedure for this achievement test battery. Appendix Table B-2 shows the raw score regression weights for the regression equations used in determining differential entry level achievement estimates, $\hat{\theta}_E$, for each subtest subsequent to the first.

Table 4
Multiple Correlations Among Ordered Subtests

Criterion		Predictor Subtest											
Subtest	Order	C	K	E	I	D	G	F	A	B	J	H	
C	1												
K	2	53											
E	3	48	51 ^a										
I	4	42	45	53									
D	5	46	49	52	57								
G	6	41	45	52	55	55							
F	7	36	38	44	51	53	53						
A	8	40	47	49	49	51	52	52					
B	9	37	40	44	46	49	51	51	52				
J	10	27	31	34	43	45	46	47	47	50			
H	11	29	32	39	40	41	44	45	45	46	46		
L	12	22	27	33	35	35	36	37	39	41	44	45	

Note. Decimal points omitted.

^a Value for $R_{E.CK}$, the multiple correlation of Subtest E, predicted from Subtests C and K.

Comparison of Adaptive and Conventional Tests

Test length. The number of items administered under both the adaptive and conventional test strategies is summarized in Table 5. Appendix Table B-3 provides the frequency distribution of number of items administered by the adaptive testing strategy for each of the twelve subtests, and Table B-4 gives this frequency distribution for all subtests combined.

Table 5
Number of Items Administered in 12 Adaptive and Conventional Subtests

Comparison of Adaptive and Conventional Subtests						
Subtest	Conventional Test	Adaptive Test				Percent Reduction ^a
		Mean	S.D.	Range		
				Min	Max	
1	15	8.73	1.86	4	13	41.8
2	24	14.12	2.90	4	20	41.2
3	17	9.87	3.38	2	17	41.9
4	22	12.57	4.60	2	22	42.9
5	19	11.55	3.58	1	18	39.2
6	13	4.70	2.10	1	12	63.8
7	18	7.44	3.21	1	15	58.7
8	10	7.07	1.71	1	10	29.3
9	10	6.44	1.72	1	9	35.6
10	23	8.42	5.54	1	22	63.4
11	12	5.52	2.97	1	12	54.0
12	18	5.41	3.20	1	15	69.9
Mean	16.75	8.49	3.06	1.67	15.42	49.3
Test Battery	201	101.84	24.08	27	153	49.3

^a Computed by the formula $100 - [(\text{Mean number of items in adaptive test} / \text{mean number of items in conventional test}) \times 100]$

The data in Table 5 show substantial reductions in test length as a result of the adaptive testing strategy. For Subtest 1, 15 items were administered by the conventional procedure while from 4 to 13 items were administered by the adaptive procedure. Fifty percent of the group answered between 7 and 10 items (see Table B-3). The mean number of items administered by the adaptive strategy in Subtest 1 was 8.73, which represents a 41.8% reduction from the number of items required by the conventional test.

Similar results were observed for the other subtests. Reduction of number of items required by the adaptive test varied from a low of 29.3% for Subtest 8 to a high of 69.9% for Subtest 12, in which a mean of 5.41 items was administered by the adaptive strategy. In Subtest 12, between 3 and 7 items were administered to 50% of the testees in the adaptive strategy as compared to 18 items for each testee in the conventional test. Subtest 12 had the highest percent reduction. In all probability, this was attributable to the increased accuracy of the test entry point from the multiple regression of the scores on the eleven prior subtests.

It is interesting to note that for Subtests 5 through 12, the minimum number of items administered by the adaptive procedure was one. Table B-3 shows that for several of these subtests, a relatively substantial number of testees was administered only one item, i.e., almost 10% of the total group for Subtests 6, 11, and 12. The minimum number of items administered by the adaptive strategy was less for tests later in the adaptive testing sequence. This probably resulted from the increased use of prior test information for determining the initial item to be administered.

Although minimum numbers of items were administered at relatively high frequencies by the adaptive strategy, the maximum numbers of items were administered to very few testees (Table B-3). For Subtests 3, 4, 8, and 11 the maximum number of items administered by the adaptive strategy was the same as that administered by the conventional test; frequencies associated with these maximums were 2, 1, 5, and 1, respectively. For the remaining eight subtests, none of the testees received the same number of items in the adaptive tests as they did in the conventional test.

The conventional test battery consisted of 201 items administered to all testees. The average number of items administered by the adaptive strategy (see Table 5) was 101.84, representing a 49.3% reduction in number of items administered. The median number of items administered was 103 (see Table B-4), indicating a slight negative skew to the distribution. Fifty percent of the testees received between 86 and 119 items in the adaptive battery, representing reductions of 57.2% to 40.8% for half of the testees. As Table B-4 shows, none of the testees required all the items in the adaptive administration. The longest adaptive battery administered required 153 items for one testee, representing a 23.9% reduction in test length; the shortest adaptive battery for one testee required only 27 items, representing a test length reduction of 86.6%.

Correlation Analysis

Table 6 shows the Pearson product-moment correlation of the Bayesian achievement level estimates ($\hat{\theta}$) for the conventional and adaptive testing strategies. Eleven of the twelve correlations were greater than .90. The highest correlations were .98 for Subtests 2 and 8; the lowest was .74 for Subtest 6.

Table 6
Correlation (r) of Bayesian Achievement Level Estimates ($\hat{\theta}$)
For the Adaptive and Conventional Testing Strategies by Subtest
and Cronbach's Alpha Coefficient for the Conventional Subtests

Subtest	No. Items	r	Cronbach's Alpha
1	15	.91	.57
2	24	.98	.69
3	17	.96	.54
4	22	.97	.65
5	19	.93	.59
6	13	.74	.44
7	18	.90	.50
8	10	.98	.56
9	10	.95	.39
10	23	.92	.61
11	12	.91	.51
12	18	.94	.40

The items contributing to the Bayesian subtest achievement level estimates in the adaptive test were a subset of those used in the conventional test. Thus, to some extent, the magnitudes of the correlations in Table 6 were a function of this part-whole relationship. This is supported by a comparison with the Alpha internal consistency estimates for the conventional subtests shown in Table 6. If there were no part-whole relationship, the correlations between the achievement level estimates would be restricted by the internal consistencies. However, all the correlations were substantially higher than the Alpha values.

If the magnitude of the correlations of the two achievement estimates were primarily determined by the part-whole relationship attributable to common items, the number of items administered in a subtest would bear a strong relationship to these correlations. This was not generally the case: One of the two highest correlations ($r=.98$) was observed for Subtest 8, which had only 10 items in the conventional test, while Subtest 9, which also had 10 items, had an $r=.95$. Although Subtest 8 had the smallest percentage reduction attributable to the adaptive administration, 20.3% (see Table 5), Subtest 9 had a 45.6% reduction; and Subtest 2 ($r=.98$) had a 41.7% reduction. Subtest 6, which had the lowest r (.74), had a 63.8% reduction attributable to adaptive testing; but the highest percent reduction (69.9%) was observed for Subtest 12, for which an $r=.94$ was observed between the adaptive and conventional achievement estimates. Thus, these data suggest that the magnitudes of the correlations shown in Table 6 were not a direct function of either the number of items in the conventional tests or the internal consistency of those tests.

Information Analysis

Termination criterion .001. The first termination criterion investigated was termination of adaptive testing when no unadministered item providing an information value greater than .001 remained in the item pool for the subtest. Using

this termination criterion with a possible 15 parameterized items in Subtest 1, the mean number of items administered in the subtest was 10.55. The smallest number of items administered was five items for six testees; the largest number administered was 15 items for one testee. Of the 165 testees 4.9% were administered 10 items.

Adaptive test mean information values $[I(\hat{\theta})]$ at intervals of estimated achievement ($\hat{\theta}$) corresponding to test termination criteria of .001 and .01 are shown in Table 7. The range of estimated achievement levels was essentially the same for both criteria, although four testees obtained $\hat{\theta}$ values in the interval 2.41 to 2.60 for the .001 case. These were outside the range of $\hat{\theta}$ values obtained in the .01 case.

For 9 of the 14 intervals in which at least 10 testees were represented for both termination criteria, no significant differences were observed in mean information values. Significantly higher mean information values were observed for the .001 termination criterion in three intervals of $\hat{\theta}$: 0.21 to 0.40, 0.41 to 0.60, and 0.61 to 0.80. For the remaining two intervals in which significant differences were observed, higher mean information was observed for the .01 termination criterion. However, the differences in mean information were small, with the largest mean difference in information .12 in the 0.21 to 0.40 interval of $\hat{\theta}$.

The strong similarity of the profiles resulting from the two termination criteria for Subtest 1 and the lack of any general trend in direction of the significant differences suggested that little was to be gained by use of the more stringent .001 termination criterion. Therefore, the remainder of the analyses were conducted with the .01 termination criterion.

Termination criterion .01. Appendix Tables B-5 through B-16 include mean raw values of estimated information $[I(\hat{\theta})]$ at intervals of $\hat{\theta}$ for the adaptive and conventional tests for ordered Subtests 1 through 12. These values are based on mean information in test items actually administered to each testee, using the testee's $\hat{\theta}$ at the termination of each subtest. Information was computed at intervals of .02 for $\hat{\theta}$ ranging from +3.0 to -3.0. The values in these tables were smoothed for plotting by the method of moving averages, averaging across three contiguous values with non-zero frequencies in order to reduce fluctuations in the mean information values resulting from differing frequencies and/or small frequencies in the intervals of $\hat{\theta}$ (McNemar, 1969, p. 8).

Figure 3 shows a plot of the smoothed information values for Subtest 1; the smoothed values for the last subtest administered, Subtest 12, are shown in Figure 4. Appendix Figures C-1 through C-10 are plots of smoothed information values for the remaining subtests. For Subtest 1 the shape of the information curve for the adaptive test, as shown in Figure 3, was very similar to that for the conventional test. The largest differences in smoothed information values occurred at $\hat{\theta}=-1.4$, where the adaptive test's smoothed information value was 2.54 and that of the conventional test was 2.47, and at $\hat{\theta}=1.3$, where the conventional test's information value was 1.70 and that of the adaptive test was 1.93.

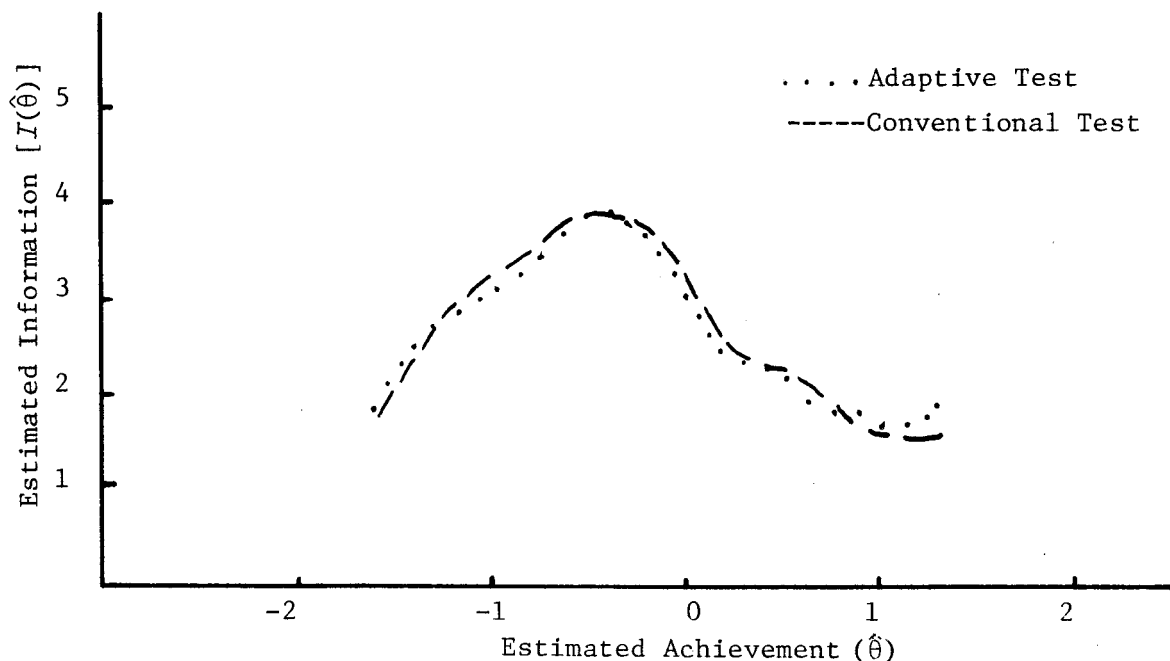
Table 7
Adaptive Test Mean Information Values [$I(\hat{\theta})$] at Estimated
Achievement Levels ($\hat{\theta}$) for Termination Criteria of .001 and .01 for Subtest 1

$\hat{\theta}$ Interval		Termination Criterion .001			Termination Criterion .01			Mean Difference [$I_{.001}(\hat{\theta}) - I_{.01}(\hat{\theta})$]	
Min	Max	N	$I(\hat{\theta})$	S.D.	N	$I(\hat{\theta})$	S.D.	t	df
-3.00	-2.80	0			0				
-2.79	-2.60	0			0				
-2.59	-2.40	0			0				
-2.39	-2.20	0			0				
-2.19	-2.00	0			0				
-1.99	-1.80	0			0				
-1.79	-1.60	10	.63	.24	11	.70	.29	-.60	19
-1.59	-1.40	19	1.72	.38	22	1.85	.40	-1.06	39
-1.39	-1.20	22	2.76	.18	21	2.87	.04	-2.74**	41
-1.19	-1.00	29	2.88	.04	23	2.86	.04	1.79	50
-0.99	-0.80	25	2.89	.07	25	2.86	.06	1.63	48
-0.79	-0.60	36	3.41	.22	33	3.36	.24	.90	67
-0.59	-0.40	21	4.19	.09	21	4.15	.15	1.05	40
-0.39	-0.20	33	4.20	.11	31	4.21	.11	-.36	62
-0.19	0.00	27	3.72	.19	27	3.72	.19	.00	52
0.01	0.20	27	3.10	.18	35	3.02	.21	1.58	60
0.21	0.40	38	2.55	.12	26	2.43	.09	4.33**	62
0.41	0.60	28	2.23	.09	42	2.17	.04	3.80**	68
0.61	0.80	23	1.97	.06	14	1.90	.00	4.34**	35
0.81	1.00	12	1.81	.05	10	1.85	.00	-2.52*	20
1.01	1.20	5	1.74	.00	13	1.74	.00		
1.21	1.40	6	1.86	.05	0				
1.41	1.60	0			11	2.19	.00		
1.61	1.80	4	2.34	.00					
1.81	2.00	0							
2.01	2.20	0							
2.21	2.40	0			0				
2.41	2.60	0			2	5.23	.32		
2.61	2.80	0			0				
2.81	3.00	0			0				

* $p < .05$ ** $p < .01$

Since mean information values were available for both adaptive and conventional tests for intervals of $\hat{\theta}$, it was possible to test the statistical significance of the difference in mean estimated information between the adaptive and conventional strategies. This was done by computing t ratios based on the raw information values in Tables B-5 through B-16 for each $\hat{\theta}$ interval containing at least ten testees in both the adaptive and conventional strategies. Computed t -ratios were based on an independent groups t -test. Although the same testees were used in determining information values for the two testing strategies, a repeated measures t -test could not be used since the same testees did not necessarily fall into the same interval of $\hat{\theta}$ on both the adaptive and conventional tests.

Figure 3
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 1

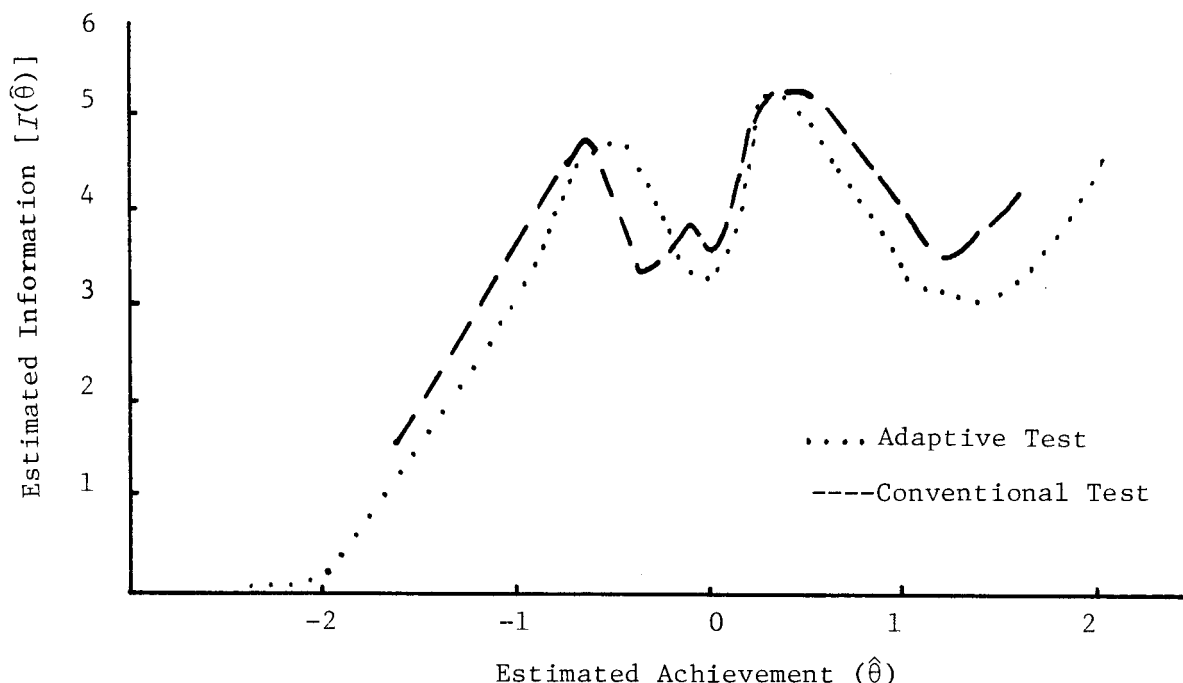


Contrasts on mean raw information values provided by the adaptive and conventional testing strategies for Subtest 1 (see Table B-5) showed significant t ratios ($p < .01$) for the $\hat{\theta}$ intervals -1.39 to -1.00 and 0.41 to 1.00 . The adaptive test provided significantly higher mean information than the conventional test over the $\hat{\theta}$ intervals -1.39 to -1.20 and 0.81 to 1.00 ; the conventional test provided significantly higher mean information than the adaptive test for the intervals -1.19 to -1.00 and 0.41 to 0.80 . For the remaining $\hat{\theta}$ intervals, there were no statistically significant differences in mean information.

Similar information curves from the two testing strategies are shown for Subtest 12 in Figure 4. Throughout the common range of $\hat{\theta}$, the two curves were very similar in shape; however, where relatively large differences in information occurred, the differences favored the conventional test. The major exception was at $\hat{\theta} = 1.5$, where the difference favored the adaptive test. For Subtest 12, the adaptive test provided $\hat{\theta}$ values in a wider range, with 46 of 365 testees obtaining $\hat{\theta}$ values less than -1.8 on the adaptive test; none of the testees obtained $\hat{\theta}$ values less than -1.8 on the conventional test.

Contrasts on mean raw information values provided by the adaptive and conventional testing strategies for Subtest 12 (see Table B-16) showed one significant t ratio ($p < .05$) for the $\hat{\theta}$ interval $-.99$ to $-.80$. In that interval the adaptive test provided significantly higher mean information than the conventional test. For the remaining $\hat{\theta}$ intervals, there were no statistically

Figure 4
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 12



significant differences between the estimated information values from the adaptive and conventional testing strategies for Subtest 12.

As shown in Tables B-6 through B-15 for Subtests 2 through 11, the overall trend was that there were few significant differences between the estimated information values at all $\hat{\theta}$ intervals where t -tests were computed. The largest number of $\hat{\theta}$ intervals for which statistically significant differences in estimated information values were obtained was 6 of a possible 14 contrasts for Subtest 1 (Table B-5); for that subtest two of the differences favored the adaptive test and four favored the conventional test. Two of the subtests (3 and 10) showed no statistically significant differences in mean estimated information values between conventional and adaptive testing. The general lack of differences in the information curves is reflected in the plots of smoothed information values for Subtests 2 through 11 shown in Appendix Figures C-1 through C-10.

Discussion

This paper has presented an adaptive testing strategy designed for use with the achievement test batteries covering multiple content areas. One goal of the strategy was to select and administer items within a subtest as a function of the amount of information provided by each item at each testee's current estimated achievement level. A second goal was to use redundant information between and among subtests, by predicting a testee's performance on subsequent

subtests based on performance on previous subtests, to determine appropriate differential entry points in adaptive branching between subtests. It was hypothesized that attaining these goals in the design of an adaptive testing strategy would result in considerable reduction in the number of items administered to each testee, while sacrificing little, if any, test information compared to that obtainable by administering the entire test battery conventionally. Thus, the focus of this adaptive testing strategy is utilization of an existing item pool for an achievement test battery to efficiently measure or estimate each testee's achievement level.

Applicability of the ICC Model

In order to implement the adaptive testing strategy, it is necessary to first obtain item parameters using the item characteristic curve (ICC) model. These parameters are then used to compute an information curve for each test item. The item information curves are used, in turn, in the process of intra-test branching.

The calibration of the achievement test items used in this study by the ICC model permits an opportunity to determine the applicability of that model to achievement test data. Bejar, Weiss, and Kingsbury (1977) specifically evaluated the applicability of the model to a college classroom achievement test. They found that 78% of the 309 items they studied yielded ICC item parameter estimates. In the present study, 87% of the items submitted to Urry's (1976) calibration procedure resulted in item parameter estimates acceptable by Urry's criteria.

Items were calibrated within content areas in the present study, while in the Bejar et al. study, calibration was in the context of the total set of items. Nevertheless, both studies showed that the achievement test items analyzed had sufficiently high discrimination parameters to be useful in adaptive testing. In the present study, the mean discrimination (a) of all the test items was 1.69; the corresponding value in the Bejar et al. study was 1.20. There was, however, a substantial difference in the c (guessing) parameter between the two studies. Although both studies used multiple-choice items with four alternative answers, the mean value of the c parameter in the Bejar et al. study was .29; the mean value obtained in the present study was .48.

There are at least two possible explanations for the higher c parameter estimates in the present study. The first, and more likely, explanation is that the c parameter is poorly estimated by Urry's program with the sample sizes and numbers of items used in the present study. As Gugel, Schmidt, and Urry (1976) show, the c parameter is very poorly estimated by Urry's calibration program for a minimum of 50 items and 500 persons. Consequently, when c parameters are estimated from data on as few as 10 items from 365 persons (as in the present study), it is likely that there is a wide discrepancy between the c parameter estimates and their true values. Thus, the high values of the c parameter observed in the present study may have resulted from inadequacies of the parameter estimation procedure.

A second possible explanation for the high c values is that some of the distractors in these four-choice items do not operate effectively as distractors. If this were the case, a testee with an "infinitely low level of θ " would be

able to eliminate one or more distractors and still randomly choose between the remaining answers. This is contrary to the concept of the testee with an "infinitely low value of θ " used to interpret the c parameter. Nevertheless, the possibility exists that if the elements of the set of distractors are not all on the same achievement dimension, high values of c may be found in real test data.

Intra-Subtest Branching

The intra-subtest item selection procedure utilized in this study is a variation of the maximum likelihood strategies of adaptive testing (see Weiss, 1974, pp. 62-66). Maximum likelihood adaptive testing strategies typically combine maximum likelihood scoring with selection of items based on maximum item information at the testee's current value of $\hat{\theta}$. The present strategy differs in that Bayesian scoring was used in place of maximum likelihood scoring; the maximum information item selection rule was used as in maximum likelihood adaptive testing.

In developing the intra-subtest branching scheme, consideration was given to using maximum likelihood procedures for scoring the items. However, given the requirement in maximum likelihood scoring of one correct item response and one incorrect item response before a $\hat{\theta}$ can be generated, it was determined to be unfeasible. Hence, the Bayesian scoring approach was used so that prior information could influence subsequent achievement level estimates with as few as one item administered.

In general, the use of maximum likelihood scoring and Bayesian scoring on the same data will not give numerically identical results. Although scores obtained from the two scoring methods are likely to be highly correlated, the Bayesian scoring method will result in scores which have a restricted range (Lord, 1976). This results from the fact that Owen's (1975) Bayesian scoring routine assumes a normal prior distribution of θ in the population; the result is θ estimates which are regressed toward the mean. The effect is a lack of θ estimates at the high and low ends of the distribution.

This restriction in range can affect the present branching strategy for testees whose true achievement levels are very high or very low. If there are items which provide information only at the extremes of the distribution (i.e., very difficult or very easy items of very high discrimination), it is possible that the regressed θ estimate from the Bayesian strategy will terminate testing too soon.

Future research should address itself to ways of eliminating the effects of regressed Bayesian θ estimates. One possible modification of the testing strategy would be to use Bayesian scoring only when a maximum-likelihood strategy is not feasible, i.e., after one item has been administered or when all items are answered correctly or incorrectly. When these conditions do not occur, maximum likelihood scoring could then be used. Another possibility would be to use a Bayesian scoring procedure throughout the adaptive test administration; at the termination of item administration within a subtest, estimated achievement scores could then be re-computed using maximum likelihood scoring. If continued testing were relevant, additional items would be administered and scored by maximum likelihood until additional items provided no further information.

Inter-Subtest Branching

The procedure for determining entry points into later tests in the adaptive sequence from the data obtained from earlier tests was based on a linear multiple regression of previously administered subtest scores. In order to implement this procedure, however, it was necessary to order the twelve subtests to obtain the relevant regression equations. The subtests were ordered by a procedure based on stepwise regression of subtest number correct scores, beginning with the highest correlation in the matrix.

Further research is necessary to determine an optimal and generalizable procedure for ordering a set of subtests for adaptive administration in an achievement test battery. The procedure used in this study may be sub-optimal for several reasons. First, it was based on subtest number correct scores, which are, in themselves, sub-optimal; thus, an ordering of subtests based on methods of subtest scoring which utilize more information about the items and/or testees might result in a correlation matrix with different values. This might yield a different ordering of subtests.

Second, the regression procedure used might lead to sub-optimal test entry points because regression estimates tend to underestimate extreme scores. When used with more optimal scoring methods (e.g., maximum likelihood scoring), this characteristic might require the administration of additional and unnecessary test items in order to mitigate the effects of inappropriate choice of initial items. Third, inappropriate ordering of tests might also result from the tendency of stepwise procedures to capitalize on characteristics of the data which are unique to a given sample. Thus, a relevant question for future research on procedures for subtest ordering is: given application of the same subtest ordering procedure, whether or not different subsamples from the same population will result in the same subtest ordering when measured by the same test battery.

The important question to be answered regarding the problem of inter-subtest branching is whether or not different test ordering procedures result in different orderings of subtests. If the answer were affirmative, the next question would be what effect ordering procedures would have both on the number of items administered and on the measurement characteristics of the resultant achievement estimates. The necessity to order subtests in a test battery for adaptive administration occurs only when all the intercorrelations among the subtests are neither zero nor 1.0. When the subtests intercorrelate zero with each other, there is no redundant information in scores on one subtest which will be useful in selecting the initial item for subsequent subtests. At the other extreme, if all subtests intercorrelate perfectly with each other, the information obtained from one is completely redundant with that obtained from any other; and no further testing is necessary.

There is one other situation in which it may not be necessary to order the subtests for adaptive administration of a test battery. This would occur when all the subtests in the battery have equal correlations with each other. In this case the multiple correlations of each subtest with every other subtest would be equal, and each subtest would provide an equal amount of redundant information.

There are other procedures for ordering subtests which need to be investigated. For example, subtests might be ordered in terms of the number of items or their reliabilities. If subtests were ordered by number of items, it would seem logical to administer the shorter tests first, based on the assumption that as differential entry points become more accurate due to additional redundant information, the longer subtests would be more useful later in the battery. When ordering tests by their reliabilities, it would seem appropriate to administer the more highly reliable subtests first: More accurate redundant information would thus be obtained for selecting entry items for later tests in the adaptive sequence. It should be noted, however, that these two criteria for subtest ordering may conflict with each other, since subtest reliabilities tend to be higher for longer tests.

All subtest ordering procedures discussed thus far result in a standard ordering of subtests for all testees. However, if the philosophy of adaptive test administration were applied to the subtest ordering problem, it would imply that the *order* of subtest administration should vary for individual testees. At this stage of research in multidimensional adaptive testing, it is not clear how such an individualized inter-subtest adaptive procedure would be implemented. It would seem that, to some extent, adaptive subtest selection would be based on the level of test information in the multivariate test space at the individual's levels of $\hat{\theta}$ upon completion of previous subtests in a battery. However, specific details for the implementation of such a procedure, as well as comparisons with alternative procedures, will have to await future research.

CONCLUSIONS

The real-data simulation study in this report has supported previous research which demonstrated that a typical achievement test can yield estimates of item difficulty and discrimination parameters useful for adaptive testing. Thus, the applicability of item characteristic curve theory to the measurement of achievement has been further corroborated.

An important concern for adaptive testing using achievement test batteries is whether or not a unidimensional model can be applied across subtests. The inter-subtest multiple correlations obtained in the present study were not considered high enough to warrant the application of a unidimensional model across subtests. Instead, a multi-content branching scheme was deemed appropriate for this achievement test battery.

The results of this study have shown that by using this achievement test battery, the amount of information extracted by adaptive testing closely approximated that for conventional testing. The number of incidences of significant differences between the information curves for the conventional and adaptive strategies was minimal, and there were no significant differences in the majority of the information values for the two testing strategies in each of the twelve subtests. Given these results, an obvious question regarding the administration of achievement test batteries is: If a computer terminal is available for test administration, why should test time be spent administering those test items which do not add to the precision of measurement on the test battery?

The adaptive testing strategy described in this report provides methods for intra-subtest and inter-subtest branching which exclude the administration of unnecessary items. The data indicate that on this achievement test battery the length of the battery can be reduced by 50% for the typical testee. In no case was it necessary to administer in the adaptive battery all of the items included in the conventional tests. Therefore, adaptive testing can reduce the time spent in testing; the time saved could then be used by the testees for other activities, such as additional instruction. It is also possible that adaptive achievement testing might have positive psychological advantages (e.g., Betz & Weiss, 1976), providing further beneficial effects on the psychometric characteristics of test scores. At the least, reduced testing time might result in more favorable attitudes of the testees toward the testing process.

In the adaptive testing strategy implemented in this study, test length is a direct function of the termination criterion employed. Testing was terminated within a subtest when none of the remaining items had a corresponding level of item information greater than .01 (.001 for Subtest 1) at the testee's current estimated achievement level; this value was arbitrarily chosen. More research is needed to determine optimal termination criteria.

That the information curves resulting from the adaptive and conventional strategies were found to be highly correspondent was to be expected from the way in which items were selected (based on item information) for the adaptive strategy. However, because of the inapplicability of maximum likelihood scoring in the early stages of item administration within a subtest, additional research is needed to develop and evaluate optimal procedures for item scoring. In addition, further research is needed for identification and evaluation of optimal procedures to order subtests for inter-subtest branching.

One additional finding from the present study was that the adaptive testing strategy consistently provided a wider range of achievement estimates than did the conventional strategy, using the same method for estimating θ . Weiss (1973) predicted that this would occur in adaptive testing. The major implication of this finding is that adaptive testing can provide more discriminating measurement in the upper and lower extremes of the achievement continuum.

This study has demonstrated that an adaptive testing strategy, designed specifically for achievement test batteries, can substantially reduce the number of items administered in all subtests of the battery without reducing the precision of subtest scores. The strategy appears to be generalizable; it should be applicable to a variety of test batteries in which there is a fixed and relatively small subset of items for each subtest. Further research is needed to evaluate the performance of this adaptive testing strategy in other test batteries and in live testing situations. In addition, research is needed to modify the adaptive testing strategy to identify optimal procedures for the complete individualized administration of an achievement test battery.

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, 1977.
- 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, 1977.
- Betz, N. E., & Weiss, D. J. An empirical study of computer-administered two-stage ability testing (Research Report 73-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD 768993)
- 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, 1975. (NTIS No. AD A013185)
- 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)
- 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, 1974. (NTIS No. AD A001230)
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968, Chapters 17-20.
- Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1970.
- Gugel, J. F., Schmidt, F. L., & Urry, V. W. Effectiveness of the ancillary correction procedure. In C. L. Clark (Ed.), Proceedings of the first conference on computerized adaptive testing. Washington, DC: U.S. Civil Service Commission, 1976.
- 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, 1975. (NTIS No. AD A006733)
- 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, 1974. (NTIS No. AD 783553)

- Lord, F. M. Discussion. In Gorham, W. A., Gugel, J. F., Lord, F. M., Jensema, C., Schmidt, F. L., & Urry, V. W. Computers and testing: Steps toward the inevitable conquest. Symposium presented at the 83rd Annual Convention of the American Psychological Association, Chicago, 1975. (PB-261-694) Washington, DC: U.S. Civil Service Commission. (NTIS No. AD 16851469).
- Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 117-138.
- 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)
- McNemar, Q. Psychological statistics. (4th Ed.) New York, NY: Wiley, 1969.
- 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.
- 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. Washington, DC: U.S. Civil Service Commission, 1976.
- 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 simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A020961)
- 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., & Betz, N. E. Ability measurement: Conventional or adaptive? (Research Report 73-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD 757788)
- 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 768376)

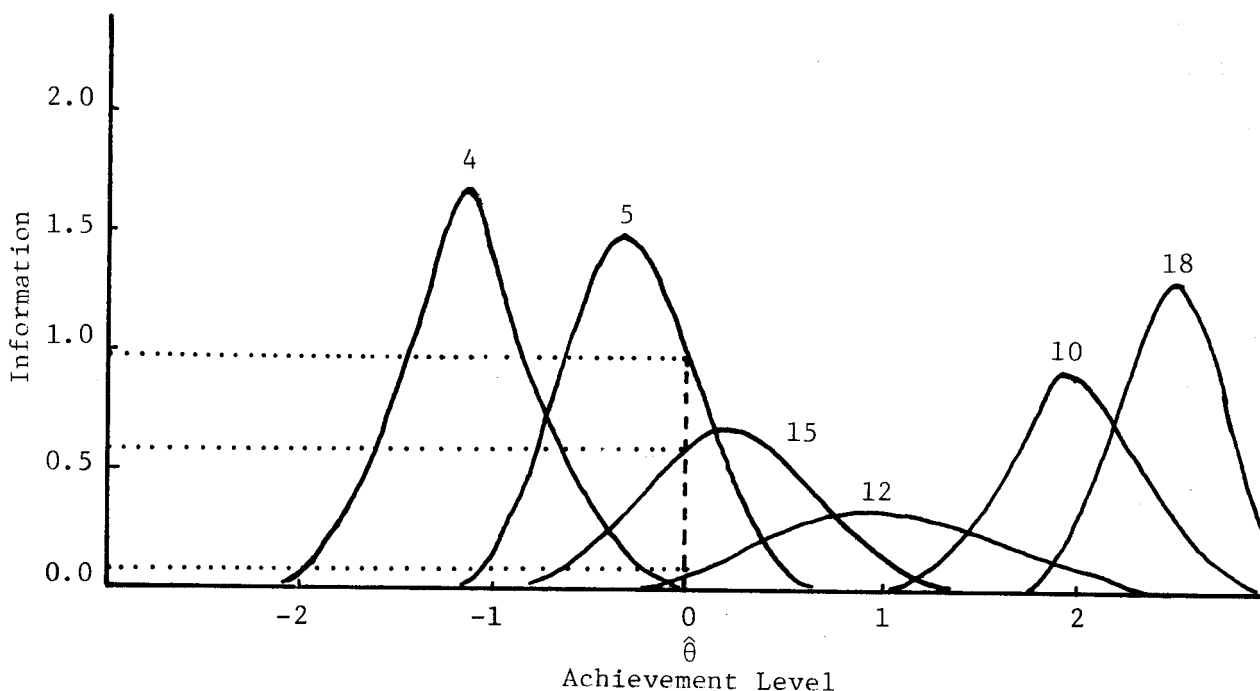
APPENDIX A

Illustration of Intra-Subtest Adaptive Branching

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 Subtest 1. (There are a total of 15 items in Subtest 1 from which only six were chosen to simplify the illustration.) The height of the information curve at a given achievement level 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.

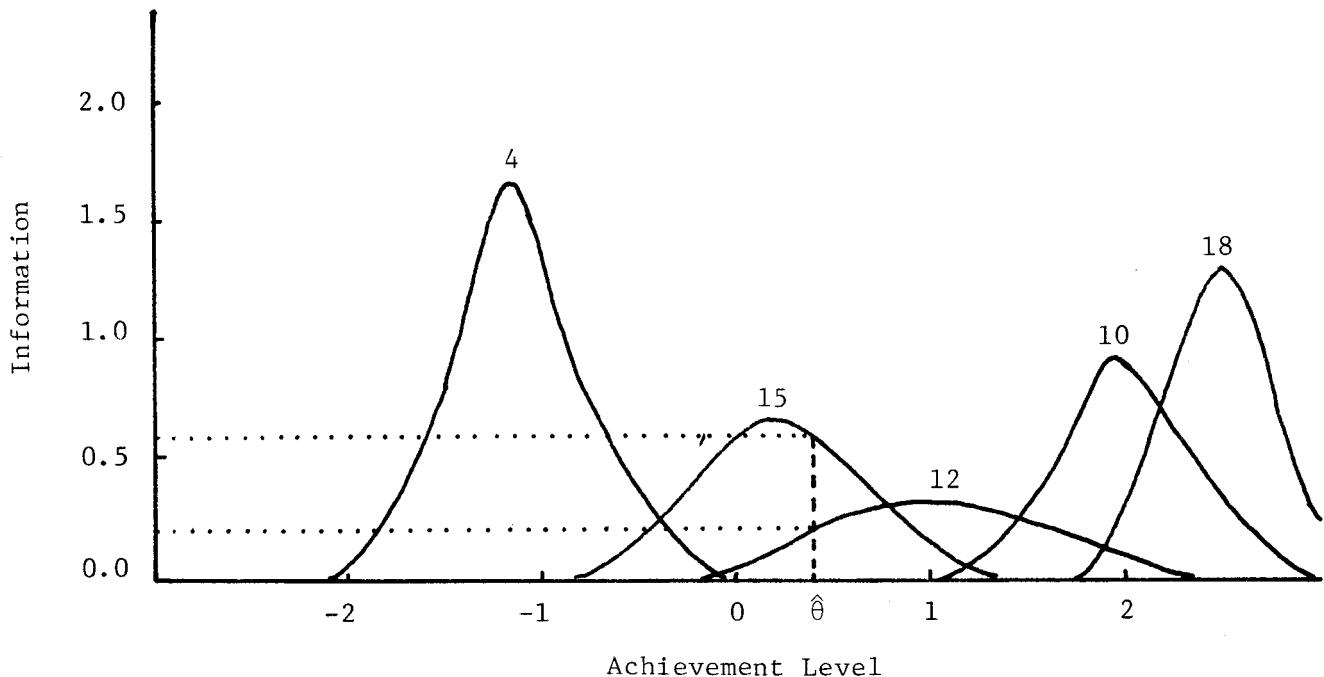
Figure A-1
Estimated Item Information Curves for Six Items from Test 1



For a testee beginning Subtest 1, the initial achievement estimate was $\hat{\theta}=0$ (this varied by individual for subsequent subtests); this is shown by the vertical dashed line in Figure A-1. Of the six items in the example, only three items had essentially non-zero information values at $\hat{\theta}=0$; these values, shown by the horizontal dotted lines in Figure A-1, were .90 for item 5, .48 for item 15, and .04 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 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 non-zero values of information. Since item 15 has an information value of .54 and item 12 has a value of .20, item 15 is 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 $\hat{\theta}$ increased to .92; this is shown in Figure A-3. At that value of $\hat{\theta}$, item 12 provides .22 information and item 10 provides .02 information. Item 12 is thus administered next. Assuming that item 12 was answered incorrectly, the $\hat{\theta}$ decreased to .62, which is plotted in Figure A-4. The figure shows that of the three items remaining, none provides any information at the current level of $\hat{\theta}$. Thus, there is no need for administering additional items from Subtest 1, and testing in that subtest is terminated. The achievement level estimate of $\hat{\theta}_1=.62$ is taken as the testee's score on Subtest 1, since it is based on all items providing more than non-trivial amounts of information about that testee's achievement level. For inter-test branching, $\hat{\theta}_1=.62$ is used in the regression equation to determine the entry point $\hat{\theta}$ estimate for selecting the first item to be administered in Subtest 2.

Figure A-3
Estimated Item Information Curves for Four Items from Test 1

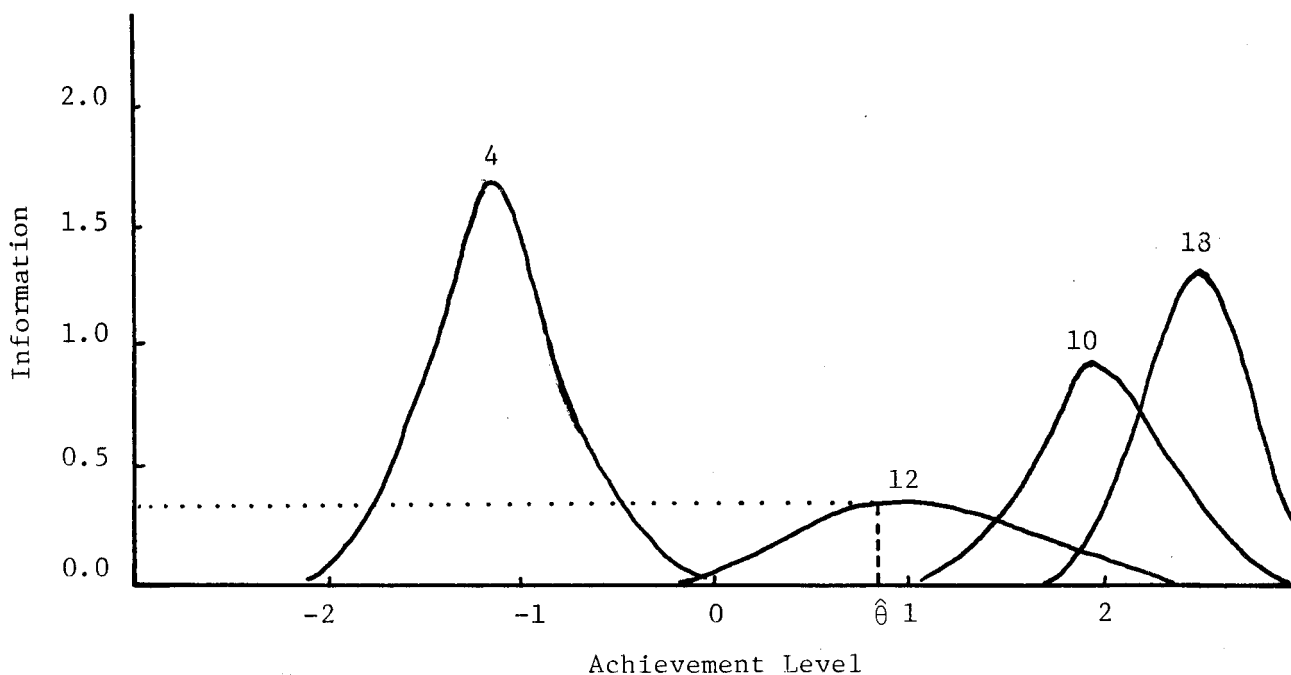
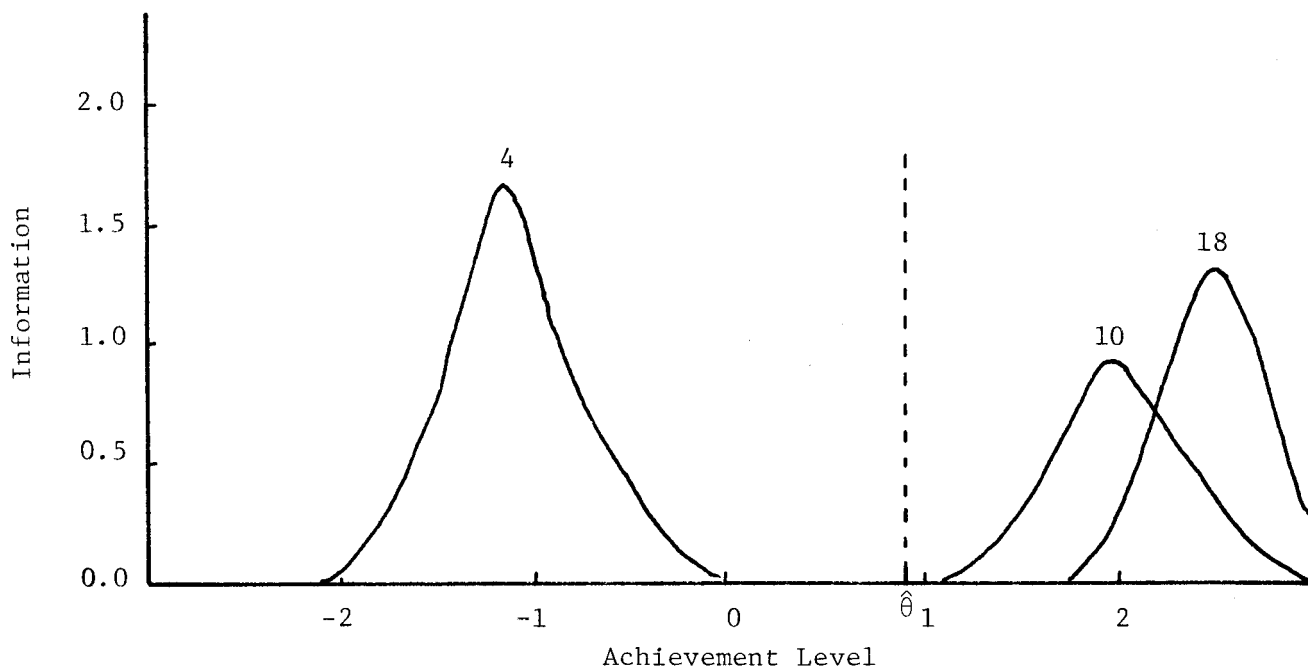


Figure A-4
Estimated Item Information Curves for Three Items from Test 1



APPENDIX B

Supplementary Tables

Table B-1
Normal Ogive Item Discrimination (*a*), Difficulty (*b*) and Guessing (*c*) Parameter
Estimates for the Twelve Subtests

Subtest and Item	<i>a</i>	<i>b</i>	<i>c</i>	Subtest and Item	<i>a</i>	<i>b</i>	<i>c</i>	Subtest and Item	<i>a</i>	<i>b</i>	<i>c</i>	Subtest and Item	<i>a</i>	<i>b</i>	<i>c</i>	Subtest and Item	<i>a</i>	<i>b</i>	<i>c</i>
Subtest A				Subtest D (continued)				Subtest F (continued)				Subtest I (continued)				Subtest K (continued)			
1	2.22	1.80	.70	8	---	---	---	14	1.97	-.41	.59	15	1.18	-.56	.31	8	---	---	---
2	1.76	-.06	.53	9	1.52	2.15	.53	15	1.00	2.78	.48	16	.90	2.00	.49	9	.97	2.75	.49
3	2.57	-1.13	.54	10	.97	-.55	.39	16	2.11	3.05	.71	17	1.65	-.82	.39	10	---	---	---
4	.88	.59	.48	11	1.69	-1.13	.48	17	1.24	2.78	.44	18	1.89	.12	.43	11	1.21	1.83	.59
5	3.01	-.52	.30	12	1.94	1.08	.52	18	1.74	2.74	.49	19	1.08	.76	.53	12	2.00	-.05	.30
6	1.29	-.51	.51	13	1.39	.15	.51	Subtest G				20	1.64	-.58	.37	13	2.36	-.66	.25
7	1.52	1.69	.65	14	---	---	---	1	3.61	-1.29	.46	21	1.23	-.97	.34	14	---	---	---
8	2.14	-.87	.54	15	1.57	-1.37	.52	2	---	---	---	22	1.69	.42	.41	15	.88	.06	.33
9	1.93	.29	.50	16	1.07	.37	.49	3	1.75	1.71	.66	23	2.00	-.97	.42	16	---	---	---
10	1.64	-.04	.47	17	---	---	---	4	1.87	1.42	.64	24	3.50	2.30	.29	17	1.67	.18	.32
Subtest B				18	2.10	-1.09	.59	5	1.19	.11	.59	Subtest J				18	---	---	---
1	3.02	1.63	.71	19	1.15	.44	.49	6	1.35	.27	.61	1	1.78	2.30	.34	19	1.37	2.80	.64
2	1.48	-.62	.36	20	1.14	1.95	.39	7	1.67	.08	.63	2	2.04	-.03	.54	20	.75	.44	.35
3	3.62	-1.65	.18	21	1.29	.55	.50	8	1.24	2.38	.47	3	1.23	.93	.61	21	1.12	-.42	.21
4	1.66	.04	.54	22	1.72	-1.34	.53	9	4.30	1.83	0.00	4	2.94	-1.29	.77	22	1.72	-1.00	.43
5	2.44	-.80	.46	Subtest E				10	1.89	.25	.64	5	---	---	---	23	.73	.49	.32
6	1.28	.24	.53	1	2.13	.62	.49	11	1.23	2.60	.37	6	1.37	.24	.45	24	1.28	2.68	.64
7	2.86	2.94	.86	2	1.20	.76	.31	12	1.84	.91	.62	7	1.32	1.97	.34	25	2.74	.34	.22
8	.90	.58	.50	3	1.05	1.78	.47	13	2.17	2.91	.49	8	1.71	-.20	.57	26	---	---	---
9	2.09	.14	.54	4	.98	.82	.49	14	1.61	2.44	.52	9	1.86	2.45	.21	27	1.66	-.18	.32
10	1.81	.64	.58	5	1.51	-.48	.38	Subtest H				10	1.13	1.80	.39	28	2.36	-.82	.38
Subtest C				6	1.42	.43	.64	1	2.00	-.08	.41	11	1.18	2.40	.33	29	1.33	3.00	.62
1	.98	2.55	.46	7	1.25	2.65	.42	2	1.87	1.38	.45	12	1.51	1.48	.39	30	---	---	---
2	---	---	---	8	---	---	---	3	3.12	-.98	.65	13	2.47	2.44	.33	31	.79	.91	.35
3	2.20	-.56	.48	9	1.59	-.71	.41	4	2.61	1.42	.37	14	1.05	1.15	.47	32	.85	2.30	.48
4	2.87	-1.37	.57	10	2.03	1.97	.59	5	3.34	-.95	.53	15	---	---	---	Subtest L			
5	2.26	-.43	.43	11	.98	1.48	.51	6	2.01	.74	.38	16	1.94	-.68	.64	1	---	---	---
6	1.68	-.73	.51	12	1.09	-.68	.37	7	2.55	.48	.36	17	1.62	1.55	.40	2	3.36	2.09	.10
7	---	---	---	13	1.98	-.64	.47	8	3.41	2.62	.54	18	---	---	---	3	---	---	---
8	---	---	---	14	3.60	2.48	.34	9	.94	2.08	.37	19	2.66	2.12	.18	4	1.29	.75	.55
9	1.35	.48	.61	15	1.36	-1.37	.46	10	.86	1.11	.32	20	1.25	.99	.59	5	1.88	-.32	.54
10	2.14	1.52	.58	16	1.78	-.33	.40	11	1.65	.78	.45	21	1.00	2.51	.37	6	1.68	2.52	.34
11	1.83	-1.28	.52	17	1.05	.75	.50	12	1.06	1.52	.35	22	1.21	1.32	.55	7	---	---	---
12	1.23	.82	.54	18	1.76	2.96	.70	Subtest I				23	1.92	1.37	.41	8	1.29	3.11	.56
13	1.06	.75	.52	Subtest F				1	1.39	1.22	.34	24	1.64	.03	.54	9	1.58	-.35	.63
14	2.17	2.50	.66	1	1.55	.18	.66	2	1.00	1.44	.25	25	.88	1.87	.37	10	2.40	-1.01	.90
15	1.51	.01	.50	2	1.30	-.35	.57	3	1.89	1.72	.43	26	---	---	---	11	1.23	2.59	.52
16	1.78	1.58	.70	3	1.42	2.43	.51	4	.77	1.15	.44	27	---	---	---	12	---	---	---
17	1.36	.21	.51	4	1.24	-.02	.55	5	---	---	---	28	---	---	---	13	---	---	---
18	2.57	2.08	.62	5	1.76	-1.07	.53	6	1.44	-.28	.35	29	2.52	2.65	.33	14	1.44	.01	.47
Subtest D				6	1.62	.53	.55	7	1.11	2.25	.51	Subtest K				15	2.41	2.23	.27
1	.94	.80	.47	7	1.78	2.61	.67	8	1.29	3.06	.57	1	1.75	-1.36	.61	16	1.67	.53	.57
2	1.40	.49	.51	8	.99	1.07	.40	9	1.04	-.32	.49	2	2.11	1.87	.46	17	2.13	2.99	.61
3	3.36	2.06	.33	9	1.27	1.98	.53	10	1.49	3.09	.64	3	---	---	---	18	.87	2.14	.45
4	1.03	2.59	.30	10	2.74	-1.05	.59	11	---	---	---	4	1.06	.17	.47	19	1.19	2.59	.54
5	1.67	-.20	.52	11	1.85	.68	.70	12	2.06	.47	.44	5	.79	1.34	.43	20	2.01	.44	.55
6	1.92	-.36	.52	12	1.24	2.75	.53	13	1.68	3.16	.59	6	2.42	2.71	.57	21	1.45	1.08	.57
7	2.51	2.26	.42	13	1.66	.82	.55	14	.89	.73	.41	7	1.55	2.44	.50	22	1.68	3.17	.59
																23	1.66	1.39	.59

Note. Dashed lines indicate that an item was rejected in the first phase of the item parameterization procedure.

Table B-2
Raw Score Regression Weights (B) for Regression Equations
Used to Determine Differential Entry Points in Inter-Subtest Branching

Ordered Subtest	Ordered Subtest											
	1	2	3	4	5	6	7	8	9	10	11	12
2	.53	.00										
3	.38	.14	3.97									
4	.29	.10	.43	4.44								
5	.27	.09	.09	.24	4.97							
6	.10	.06	.20	.10	.06	.85						
7	.08	.02	.12	.18	.12	.12	2.38					
8	.09	.10	.10	-.05	.09	.04	.08	1.55				
9	.05	.02	.07	.02	.10	.13	.00	.08	1.70			
10	-.03	.05	-.01	.20	.09	.16	.16	-.03	.36	4.05		
11	.16	.02	.10	.00	.03	.16	.09	.06	.10	.06	.25	
12	-.01	-.04	.09	.08	-.13	.07	-.07	.23	.14	.15	.14	6.13

Note. Regression constants (A) are on the main diagonal.

Table B-3
Frequency of Number of Items Administered by the Adaptive Testing
Strategy and Number of Items in the Conventional Subtest (*) for
Each of the Twelve Subtests (N=365 Testees)

No. of Items Administered	Subtest											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	1	37	9	11	6	8	33	30
2	0	0	2	12	6	21	30	3	5	36	17	44
3	0	0	8	7	9	41	12	4	37	29	70	40
4	10	2	11	7	2	39	19	3	14	28	45	67
5	29	8	9	6	9	120	38	41	7	18	37	27
6	8	5	73	7	13	43	38	29	22	73	31	29
7	108	0	20	19	23	36	21	62	191	28	18	33
8	127	14	10	28	24	18	24	186	82	8	21	25
9	18	6	7	18	15	4	55	21	1	8	58	35
10	8	12	10	16	13	2	83	5*	0*	16	17	11
11	2	12	75	13	15	1	10			11	17	10
12	0	2	42	11	13	3	10			13	1*	1
13	0*	33	58	26	81	0*	5			4		3
14		24	26	29	85		6			19		7
15		147	7	51	38		5			16		3
16		55	5	36	15		0			4		0
17		32	2*	35	2		0			9		0
18		10		31	1		0*			11		0*
19		1		11	0*					6		
20		2		0						11		
21		0		1						8		
22		0		1*						1		
23		0								0*		
24		0*										

Note. 25th and 75th percentiles are underlined.

Table B-4
Frequency and Cumulative Percent of Total Number of Items Administered by
the Adaptive Testing Strategy Across all 12 Subtests (N=365 Testees)

No. Items	Freq.	Cum Pct.	No. Items	Freq.	Cum Pct.	No. Items	Freq.	Cum Pct.
27	1	1	85	8	24	117	1	72
41	1	1	86	6	25	118	5	73
42	1	1	87	1	25	119	8	75
44	3	2	88	5	27	120	3	76
47	3	2	89	4	28	121	3	77
51	3	3	90	8	30	122	2	78
52	3	4	91	2	31	123	4	79
54	1	4	92	4	32	124	6	80
55	2	5	93	5	33	125	4	81
57	2	5	94	6	35	126	4	82
58	1	6	95	7	37	127	9	85
59	3	7	96	5	38	128	6	87
60	2	7	97	4	39	129	5	88
61	2	8	98	8	41	130	5	89
65	3	8	99	7	43	131	2	90
66	4	10	100	5	45	132	4	91
68	1	10	101	8	47	133	4	92
69	2	10	102	4	48	134	2	93
70	4	12	103	6	50	135	3	93
71	2	12	104	4	51	136	3	94
72	1	12	105	6	52	137	1	95
73	1	13	106	8	55	138	2	95
74	4	14	107	2	55	139	5	96
75	2	14	108	11	58	141	2	97
76	5	16	109	11	61	142	2	98
77	1	16	110	6	63	144	2	98
78	3	17	111	6	64	145	2	99
79	2	17	112	8	67	146	1	99
80	5	19	113	6	68	147	1	99
81	1	19	114	4	69	148	1	99
82	2	19	115	4	70	149	1	99
83	4	21	116	4	72	153	1	100
84	3	21						

Table B-5
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$
and Mean Difference in Information and t Values
at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 1

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	SD	N	$I_c(\hat{\theta})$	SD	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	0			0					
-2.39	-2.20	0			0					
-2.19	-2.00	0			0					
-1.99	-1.80	0			0					
-1.79	-1.60	11	.70	.29	14	.64	.23	-.06	-.58	23
-1.59	-1.40	22	1.85	.40	23	1.83	.35	-.02	-.18	43
-1.39	-1.20	21	2.87	.04	25	2.73	.18	-.14	-3.49**	44
-1.19	-1.00	23	2.86	.04	20	2.89	.03	.03	2.75**	41
-0.99	-0.80	25	2.86	.06	28	2.89	.06	.03	1.82	51
-0.79	-0.60	33	3.36	.24	37	3.38	.19	.02	.39	68
-0.59	-0.40	21	4.15	.15	19	4.15	.16	.00	.00	38
-0.39	-0.20	31	4.21	.11	24	4.26	.06	.05	2.01	53
-0.19	0.00	27	3.72	.19	32	3.75	.23	.03	.54	57
0.01	0.20	35	3.02	.21	30	3.04	.21	.02	.38	63
0.21	0.40	26	2.43	.09	31	2.50	.12	.07	2.45*	55
0.41	0.60	42	2.17	.04	29	2.23	.08	.06	4.17**	69
0.61	0.80	14	1.90	.00	27	1.96	.07	.06	3.19**	39
0.81	1.00	10	1.85	.00	10	1.81	.04	-.04	-3.16**	18
1.01	1.20	13	1.74	.00	7	1.74	.01	.00		
1.21	1.40	0			6	1.85	.01			
1.41	1.60	11	2.10	.00	3	2.13	.00	-.06		
1.61	1.80				0					
1.81	2.00				0					
2.01	2.20				0					
2.21	2.40				0					
2.41	2.60				0					
2.61	2.80				0					
2.81	3.00				0					

* $p \leq .05$

** $p \leq .01$

Table B-6
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 2

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	2	.00	.00	0					
-2.39	-2.20	16	.00	.00	0					
-2.19	-2.00	20	.01	.01	0					
-1.99	-1.80	9	.03	.02	0					
-1.79	-1.60	7	.16	.16	58	.32	.07	.16		
-1.59	-1.40	1	1.85		0					
-1.39	-1.20	20	3.50	.37	0					
-1.19	-1.00	40	3.14	.64	0					
-0.99	-0.80	40	1.60	.42	12	1.32	.20	-.28	-2.22*	50
-0.79	-0.60	29	.76	.36	30	.68	.15	-.08	-1.12	57
-0.59	-0.40	31	.45	.01	49	.44	.07	-.01	-.79	78
-0.39	-0.20	33	.58	.08	58	.63	.07	.05	3.11**	89
-0.19	0.00	50	.42	.53	80	.36	.46	-.06	-.68	128
0.01	0.20	11	1.44	.12	12	1.24	.40	-.20	-1.59	21
0.21	0.40	16	1.82	.05	13	1.81	.06	-.01	-.49	27
0.41	0.60	15	1.91	.01	23	1.91	.01	.00	.00	36
0.61	0.80	5	1.88	.01	6	1.88	.02	.00		
0.81	1.00	2	1.82	.01	1	1.84		.02		
1.01	1.20	0			0					
1.21	1.40	2	1.73	2.44	4	2.64	.41	.91		
1.41	1.60	1	.00		12	6.76	1.50	6.76		
1.61	1.80	1	.00		3	11.56	3.08	11.56		
1.81	2.00	1	.00		1	15.13		15.13		
2.01	2.20	0			2	8.58	2.82			
2.21	2.40	0			1	3.44				
2.41	2.60	1	2.55		0					
2.61	2.80	0			0					
2.81	3.00	0			0					

* $p < .05$

** $p < .01$

Table B-7

Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels ($\hat{\theta}$) for Subtest 3

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0								
-2.79	-2.60	0								
-2.59	-2.40	0								
-2.39	-2.20	12	.00	.00						
-2.19	-2.00	10	.02	.01						
-1.99	-1.80	27	.04	.02						
-1.79	-1.60	17	.14	.05	1	.25				
-1.59	-1.40	19	.49	.17	12	.48	.13	-.01	-.17	29
-1.39	-1.20	17	1.03	.24	32	1.14	.24	.11	1.53	47
-1.19	-1.00	36	2.05	.28	31	1.97	.27	-.08	-1.19	65
-0.99	-0.80	15	2.37	.66	28	2.59	.05	.22	1.77	41
-0.79	-0.60	42	2.47	.06	40	2.44	.08	-.03	-1.93	80
-0.59	-0.40	21	2.12	.49	30	2.22	.04	.10	1.12	49
-0.39	-0.20	26	2.16	.01	33	2.16	.01	.00	.00	57
-0.19	0.00	42	.66	.99	79	.89	1.06	.23	1.16	119
0.01	0.20	15	2.00	.55	26	1.98	.58	-.02	-.11	39
0.21	0.40	9	2.28	.05	18	2.28	.05	.00	.00	25
0.41	0.60	16	2.52	.05	15	2.48	.06	-.04	-2.02	29
0.61	0.80	4	2.60	.02	5	2.65	.02	.05		
0.81	1.00	5	2.08	1.16	6	2.55	.07	.47		
1.01	1.20	7	2.29	.12	3	2.28	.06	-.01		
1.21	1.40	6	1.86	.11	3	1.83	.09	-.03		
1.41	1.60	5	1.22	.69	0					
1.61	1.80	1	.00		1	.00		.00		
1.81	2.00	1	.00		0					
2.01	2.20	0			0					
2.21	2.40	0			0					
2.41	2.60	0			0					
2.61	2.80	0			0					
2.81	3.00	1	.00		0					

Table B-8
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 4

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0								
-2.79	-2.60	0								
-2.59	-2.40	0								
-2.39	-2.20	1	.00							
-2.19	-2.00	6	.01	.00						
-1.99	-1.80	17	.03	.02						
-1.79	-1.60	12	.17	.07	3	.21	.06	.04		
-1.59	-1.40	18	.44	.13	12	.49	.16	.05	.94	28
-1.39	-1.20	25	1.37	.37	20	1.38	.32	.01	.10	43
-1.19	-1.00	25	2.59	.37	29	2.61	.31	.02	.22	52
-0.99	-0.80	14	3.87	.41	24	3.63	.36	-.24	-1.88	36
-0.79	-0.60	23	5.08	1.26	33	5.06	.42	-.02	-.09	54
-0.59	-0.40	22	6.31	.11	24	6.30	.08	-.01	-.36	44
-0.39	-0.20	21	5.39	1.30	17	5.70	.32	.31	.96	36
-0.19	0.00	57	1.72	2.04	81	1.58	1.99	-.14	-.40	136
0.01	0.20	22	3.21	.24	33	3.17	.59	-.04	-.30	53
0.21	0.40	6	2.58	.10	10	2.49	.00	-.09		
0.41	0.60	29	2.19	.14	23	2.20	.14	.01	.26	50
0.61	0.80	20	1.57	.14	23	1.71	.15	.14	3.15**	41
0.81	1.00	15	1.31	.09	14	1.41	.01	.10	4.13**	27
1.01	1.20	5	.84	.04	18	1.00	.00	.16		
1.21	1.40	10	.78	.01	0					
1.41	1.60	3	.85	.07	0					
1.61	1.80	1	.00		0					
1.81	2.00	2	.00	.00	0					
2.01	2.20	1	.00		0					
2.21	2.40	0			0					
2.41	2.60	0			0					
2.61	2.80	0			0					
2.81	3.00	0			0					

** $p < .01$

Table B-9
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 5

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0								
-2.79	-2.60	0								
-2.59	-2.40	6	.01	.00						
-2.39	-2.20	4	.07	.05						
-2.19	-2.00	9	.27	.09						
-1.99	-1.80	7	3.31	.79	21	3.48	.69	.17		
-1.79	-1.60	6	5.39	2.66	10	4.98	.38	-.41		
-1.59	-1.40	8	5.64	.62	0					
-1.39	-1.20	13	3.50	.62	8	2.56	.24	-.94		
-1.19	-1.00	26	2.04	.11	18	2.02	.11	-.02	-.59	42
-0.99	-0.80	38	2.22	.14	25	2.19	.16	-.03	-.79	61
-0.79	-0.60	25	2.61	.07	33	2.64	.06	.03	1.76	56
-0.59	-0.40	33	2.50	.45	29	2.59	.08	.09	1.06	60
-0.39	-0.20	34	2.40	.03	31	2.40	.03	.00	.00	63
-0.19	0.00	60	1.19	.06	87	1.20	1.26	.01	.06	145
0.01	0.20	21	2.77	.08	25	2.15	1.10	-.62	-2.57**	44
0.21	0.40	14	2.68	.77	27	2.89	.01	.21	1.31	39
0.41	0.60	10	2.48	.80	16	2.77	.09	.29	1.45	24
0.61	0.80	17	2.39	.10	16	2.14	.01	-.25	-9.94**	16
0.81	1.00	2	1.89	.13	0					
1.01	1.20	6	1.34	.07	18	1.59	.02	.25		
1.21	1.40	8	1.20	.01	0					
1.41	1.60	4	1.27	.03	0					
1.61	1.80	0			0					
1.81	2.00	1	.00		0					
2.01	2.20	1	.00		1	.00		.00		
2.21	2.40	1	.00		0					
2.41	2.60	0			0					
2.61	2.80	0			0					
2.81	3.00	1	.00		0					

** $p < .01$

Table B-10

Adaptive and Conventional Test Mean Information Values [$I(\hat{\theta})$] and Mean Difference in Information and t Values at Estimated Achievement Levels ($\hat{\theta}$) for Subtest 6

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	4	.00	.00	0					
-2.39	-2.20	10	.00	.00	0					
-2.19	-2.00	10	.00	.00	0					
-1.99	-1.80	21	.02	.01	0					
-1.79	-1.60	21	.11	.03	4	.15	.03	.04		
-1.59	-1.40	19	.39	.13	32	.38	.11	-.01	-.29	49
-1.39	-1.20	35	.83	.11	6	.94	.16	.11		
-1.19	-1.00	31	1.10	.02	26	1.11	.01	.01	2.32*	55
-0.99	-0.80	16	1.08	.01	32	1.09	.01	.01	3.27**	46
-0.79	-0.60	16	1.21	.05	26	1.26	.06	.05	2.79**	40
-0.59	-0.40	43	1.55	.11	42	1.57	.15	.02	.70	83
-0.39	-0.20	8	1.78	.74	35	2.10	.19	.32		
-0.19	0.00	52	1.08	1.22	75	.99	1.33	-.09	-.39	125
0.01	0.20	18	2.73	.99	20	2.93	.69	.20	.73	36
0.21	0.40	11	3.01	.07	26	3.01	.07	.00	.00	35
0.41	0.60	10	2.51	.89	14	2.59	.75	.08	.24	22
0.61	0.80	11	2.65	.03	10	2.67	.03	.02	1.53	19
0.81	1.00	4	2.96	.18	7	2.94	.15	-.02		
1.01	1.20	4	3.60	.37	1	3.52		-.08		
1.21	1.40	2	4.57	.63	3	4.55	.41	-.02		
1.41	1.60	1	5.70		4	5.78	.31	.08		
1.61	1.80	3	6.54	.25	0					
1.81	2.00	2	8.10	.29	0					
2.01	2.20	1	.00		0					
2.21	2.40	1	.00		0					
2.41	2.60	1	.00		0					
2.61	2.80	0			1	10.57				
2.81	3.00	0			0					

* $p < .05$

** $p < .01$

Table B-11
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference
in Information and t Values at Estimated Achievement Levels ($\hat{\theta}$) for Subtest 7

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	0			0					
-2.39	-2.20	9	.00	.00	0					
-2.19	-2.00	31	.00	.00	0					
-1.99	-1.80	18	.00	.00	0					
-1.79	-1.60	6	.01	.01	0					
-1.59	-1.40	19	.09	.07	14	.09	.01	.00	.00	31
-1.39	-1.20	15	.80	.29	35	.62	.20	-.18	-2.54*	48
-1.19	-1.00	34	2.42	.71	38	2.14	.39	-.28	-2.10	70
-0.99	-0.80	26	4.15	.21	0					
-0.79	-0.60	24	3.15	.40	35	3.42	.39	.27	2.59*	57
-0.59	-0.40	47	2.12	.29	32	2.13	.33	.01	.14	77
-0.39	-0.20	17	1.55	.04	43	1.55	.04	.00	.00	58
-0.19	0.00	40	1.00	4.01	90	.78	.86	-.22	-.50	128
0.01	0.20	9	2.28	.19	21	2.17	.57	-.11	-.56	28
0.21	0.40	16	2.95	.85	11	3.33	.32	.38	1.41	25
0.41	0.60	10	4.53	.26	16	4.56	.20	.03	.33	24
0.61	0.80	8	4.32	1.74	12	4.94	.03	.62		
0.81	1.00	9	4.22	1.59	6	4.67	.03	.45		
1.01	1.20	2	4.67	.03	4	4.66	.04	-.01		
1.21	1.40	2	5.23	.02	2	5.27	.10	.04		
1.41	1.60	5	5.28	.10	2	5.29	.10	.01		
1.61	1.80	1	4.30		0					
1.81	2.00	2	2.96	.61	2	3.55	.00	.59		
2.01	2.20	1	2.40		1	1.87				
2.21	2.40	1	.00		0					
2.41	2.60	1	.00		0					
2.61	2.80	1	.00		0					
2.81	3.00	1	2.19		0					

* $p < .05$

Table B-12

Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels ($\hat{\theta}$) for Subtest 8

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	0			0					
-2.39	-2.20	8	.00	.00	0					
-2.19	-2.00	9	.00	.00	1	.00				
-1.99	-1.80	19	.00	.00	0					
-1.79	-1.60	32	.01	.01	0					
-1.59	-1.40	61	.03	.01	4	.05	.01	.02		
-1.39	-1.20	17	.11	.04	26	.14	.03	.03	2.81**	41
-1.19	-1.00	38	.31	.05	29	.29	.07	-.02	-1.36	65
-0.99	-0.80	26	.54	.08	43	.54	.08	.00	.00	67
-0.79	-0.60	25	.89	.09	59	.84	.10	-.05	-2.16	82
-0.59	-0.40	10	1.26	.16	34	1.26	.15	.00	.00	42
-0.39	-0.20	18	1.76	.12	37	1.72	.13	-.04	-1.10	53
-0.19	0.00	41	.58	.92	70	.63	.97	.05	.27	109
0.01	0.20	10	2.08	.73	26	2.24	.46	.16	.79	34
0.21	0.40	13	2.56	.07	10	2.67	.08	.11	3.51**	21
0.41	0.60	3	1.88	1.62	11	2.83	.04	.95		
0.61	0.80	7	2.78	.07	2	2.82	.06	.04		
0.81	1.00	6	1.74	1.35	6	2.58	.10	.84		
1.01	1.20	3	2.27	.06	4	2.35	.05	.08		
1.21	1.40	0			2	2.18	.01			
1.41	1.60	2	2.23	.16	1	2.18		.05		
1.61	1.80	1	4.35		0					
1.81	2.00	1	.00		0					
2.01	2.20	1	9.71		0					
2.21	2.40	0			0					
2.41	2.60	1	.00		0					
2.61	2.80	1	.00		0					
2.81	3.00	1	.00		0					

** $p < .01$

Table B-13
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 9

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0								
-2.79	-2.60	0								
-2.59	-2.40	0								
-2.39	-2.20	0								
-2.19	-2.00	1	.00							
-1.99	-1.80	0								
-1.79	-1.60	12	.64	.34	14	.64	.23	.00	.00	24
-1.59	-1.40	22	1.85	.40	23	1.83	.35	-.02	-.18	43
-1.39	-1.20	20	2.86	.04	24	2.73	.18	-.13	-3.16**	42
-1.19	-1.00	24	2.74	.59	20	2.89	.03	.15	1.13	42
-0.99	-0.80	25	2.86	.06	28	2.89	.06	.03	1.82	51
-0.79	-0.60	31	3.35	.24	34	3.39	.20	.04	.73	63
-0.59	-0.40	20	4.14	.15	18	4.14	.16	.00	.00	36
-0.39	-0.20	29	4.21	.11	22	4.25	.07	.04	1.49	49
-0.19	0.00	50	1.67	1.87	75	1.36	1.83	-.31	-.92	123
0.01	0.20	30	3.03	.21	27	2.90	.62	-.13	-1.08	55
0.21	0.40	22	2.32	.53	24	2.50	.12	.18	1.62	44
0.41	0.60	33	2.17	.04	23	2.23	.08	.06	3.70**	54
0.61	0.80	9	1.69	.63	19	1.97	.07	.28		
0.81	1.00	8	1.85	.00	5	1.82	.05	-.03		
1.01	1.20	7	1.49	.66	4	1.74	.00	.25		
1.21	1.40	0			2	1.85	.02			
1.41	1.60	6	2.19	.00	2	2.13	.00	-.06		
1.61	1.80	0								
1.81	2.00	2	.00	.00						
2.01	2.20	0								
2.21	2.40	0								
2.41	2.60	0								
2.61	2.80	1	.00							
2.81	3.00	2	.00	.00						

** $p < .01$

Table B-14

Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 10

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	0			0					
-2.39	-2.20	1	.00		0					
-2.19	-2.00	0			0					
-1.99	-1.80	11	.21	.08	3	.30	.03	.09		
-1.79	-1.60	11	.51	.15	11	.47	.09	-.04	-.76	20
-1.59	-1.40	15	.87	.28	8	1.18	.17	.31		
-1.39	-1.20	21	2.07	.32	16	1.89	.29	-.18	-1.76	35
-1.19	-1.00	15	3.39	.54	19	3.25	.51	-.14	-.78	32
-0.99	-0.80	21	4.84	1.20	28	5.23	.66	.39	1.45	47
-0.79	-0.60	24	6.79	.29	21	6.94	.25	.15	1.85	43
-0.59	-0.40	26	7.25	.04	29	7.24	.04	-.01	-.93	53
-0.39	-0.20	30	7.09	.04	32	7.07	.04	-.02	-1.97	60
-0.19	0.00	65	4.02	3.64	83	3.24	3.64	-.78	-1.29	146
0.01	0.20	22	8.05	.27	28	7.43	2.12	-.62	-1.36	48
0.21	0.40	21	8.56	.05	23	8.59	.06	.03	1.79	42
0.41	0.60	15	7.02	1.98	14	7.68	.44	.66	1.22	27
0.61	0.80	20	5.61	.65	18	5.66	.64	.05	.24	36
0.81	1.00	10	3.89	.48	15	3.88	.46	-.01	-.05	23
0.01	1.20	11	2.69	.20	11	2.09	1.05	-.60	-1.86	20
1.21	1.40	10	1.94	.69	3	2.14	.06	.20		
1.41	1.60	2	2.14	.11	3	2.08	.01	-.06		
1.61	1.80	1	.30							
1.81	2.00	0								
2.01	2.20	0								
2.21	2.40	0								
2.41	2.60	0								
2.61	2.80	0								
2.81	3.00	3	.00	.00						

Table B-15
Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference
in Information and t Values at Estimated Achievement Levels $(\hat{\theta})$ for Subtest 11

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	2	.04	.00	0					
-2.39	-2.20	4	.03	.03	0					
-2.19	-2.00	16	.13	.03	0					
-1.99	-1.80	11	.25	.05	0					
-1.79	-1.60	14	.44	.15	0					
-1.59	-1.40	21	.74	.09	6	.76	.08	.02		
-1.39	-1.20	22	1.67	.29	18	1.13	.12	-.54	-7.39**	38
-1.19	-1.00	36	1.69	.22	31	1.91	.13	.22	4.88**	65
-0.99	-0.80	22	2.61	.28	34	2.69	.30	.08	1.00	54
-0.79	-0.60	25	3.40	.74	52	3.57	.27	.17	1.47	75
-0.59	-0.40	33	4.15	.12	32	4.23	.09	.08	3.03**	63
-0.39	-0.20	25	4.27	.06	50	4.28	.06	.01	.68	73
-0.19	0.00	54	1.89	1.95	63	1.10	1.75	-.79	-2.31*	115
0.01	0.20	19	3.30	.59	32	3.09	1.02	-.21	-.82	49
0.21	0.40	12	3.19	.01	16	3.20	.02	.01	1.59	26
0.41	0.60	11	3.30	.05	11	3.27	.05	-.03	-1.41	20
0.61	0.80	8	2.95	1.19	11	3.39	.02	.44		
0.81	1.00	9	3.18	.09	1	3.33		.15		
1.01	1.20	1	.00		2	2.67	.04	2.67		
1.21	1.40	3	2.13	.03	2	2.28	.12	.15		
1.41	1.60	2	1.98	.01	1	2.00		.02		
1.61	1.80	1	2.00		0					
1.81	2.00	0			1	2.06				
2.01	2.20	1	2.71		0					
2.21	2.40	0			0					
2.41	2.60	0			0					
2.61	2.80	0			0					
2.81	3.00	0			0					

* $p < .05$

** $p < .01$

Table B-16

Adaptive and Conventional Test Mean Information Values $[I(\hat{\theta})]$ and Mean Difference in Information and t Values at Estimated Achievement Levels ($\hat{\theta}$) for Subtest 11

$\hat{\theta}$ Interval		Adaptive Test			Conventional Test			Mean Difference		
Min	Max	N	$I_a(\hat{\theta})$	S.D.	N	$I_c(\hat{\theta})$	S.D.	$[I_c(\hat{\theta}) - I_a(\hat{\theta})]$	t	df
-3.00	-2.80	0			0					
-2.79	-2.60	0			0					
-2.59	-2.40	11	.11	.32	0					
-2.39	-2.20	7	.04	.01	0					
-2.19	-2.00	15	.06	.03	0					
-1.99	-1.80	13	.20	.04	0					
-1.79	-1.60	12	.41	.07	1	.53				
-1.59	-1.40	15	.88	.28	10	.95	.17	.07	.71	23
-1.39	-1.20	23	1.73	.24	21	1.81	.26	.08	1.06	42
-1.19	-1.00	23	2.63	.67	24	2.81	.32	.18	1.18	45
-0.99	-0.80	17	4.04	.24	31	3.86	.28	-.18	-2.24*	46
-0.79	-0.60	27	4.57	.14	32	4.56	.12	-.01	-.30	57
-0.59	-0.40	33	4.64	.83	44	4.80	.01	.16	1.28	75
-0.39	-0.20	23	4.75	.02	35	4.76	.02	.01	1.86	56
-0.19	0.00	49	2.07	2.37	80	2.03	2.37	-.04	-.09	127
0.01	0.20	19	4.97	.09	24	4.36	1.69	-.61	-1.57	41
0.21	0.40	16	5.23	.05	16	5.23	.06	.00	.00	30
0.41	0.60	10	5.25	.05	12	5.27	.04	.02	1.04	20
0.61	0.80	11	4.84	.17	10	4.89	.15	.05	.71	19
0.81	1.00	10	3.35	1.77	9	4.29	.16	.94		
1.01	1.20	4	3.73	.04	7	3.60	.09	-.13		
1.21	1.40	7	2.89	1.28	5	3.36	.05	.47		
1.41	1.60	4	3.30	.00	1	3.32		.02		
1.61	1.80	1	3.32		0					
1.81	2.00	1	3.69		0					
2.01	2.20	0			0					
2.21	2.40	0			1	6.66				
2.41	2.60	1	6.67		0					
2.61	2.80	0			0					
2.81	3.00	0			1	2.26				

* $p < .05$

APPENDIX C

Supplementary Figures

Figure C-1
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 2

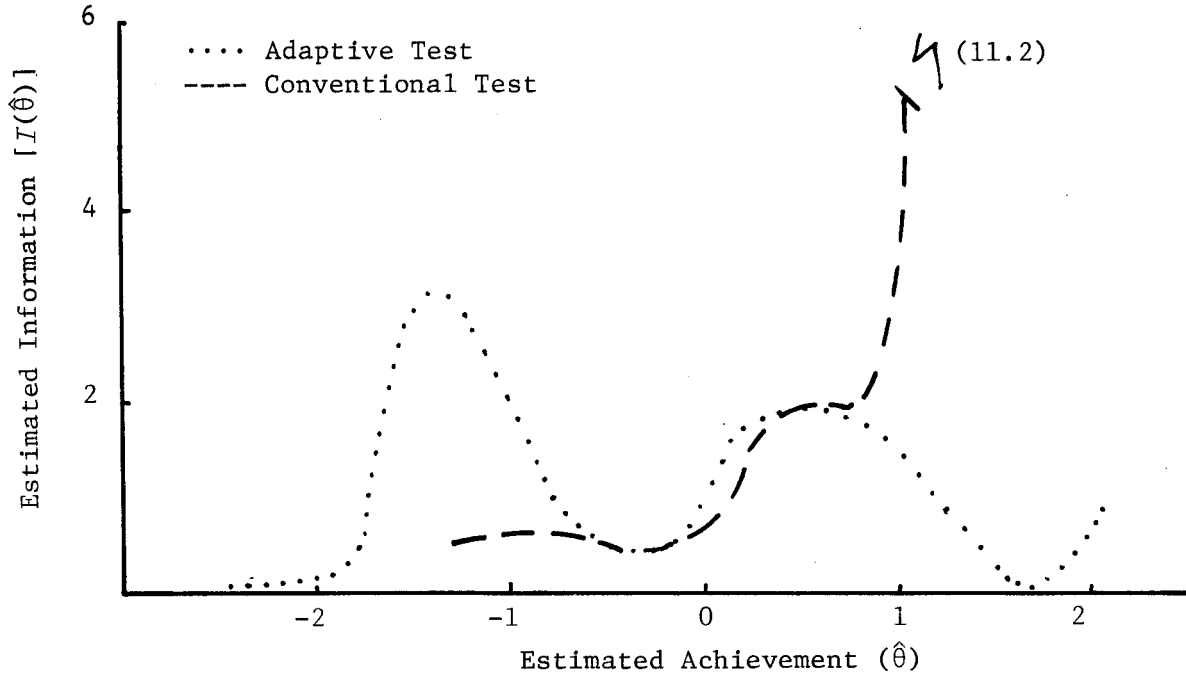


Figure C-2
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 3

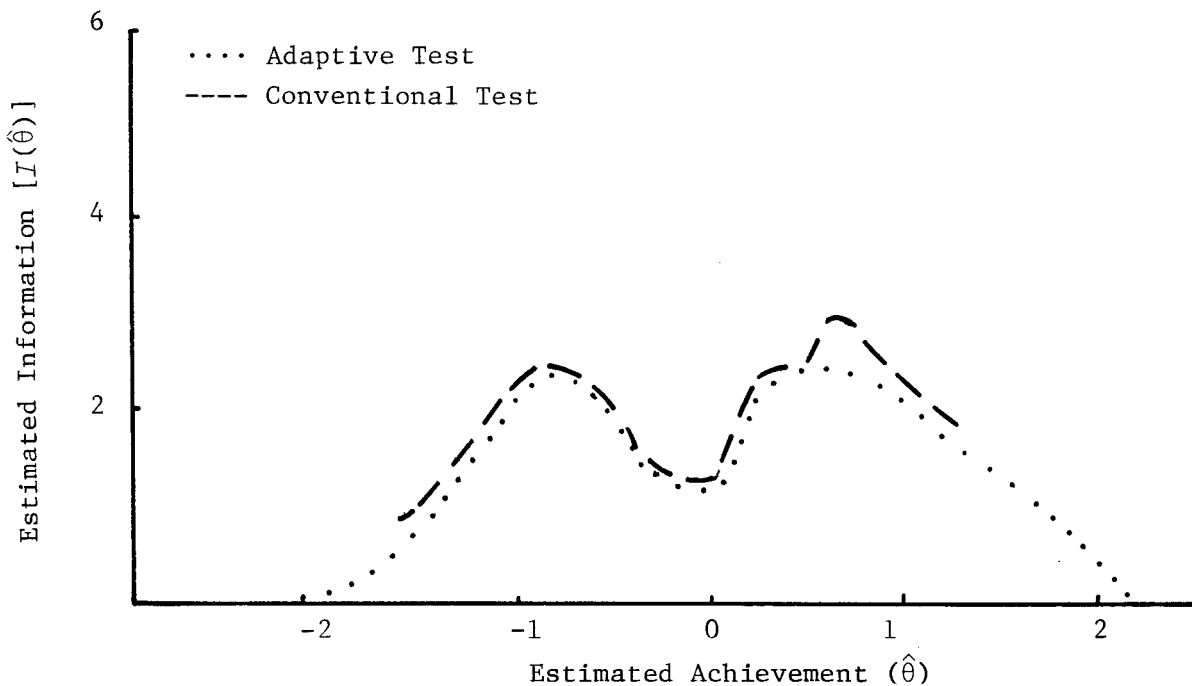


Figure C-3
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 4

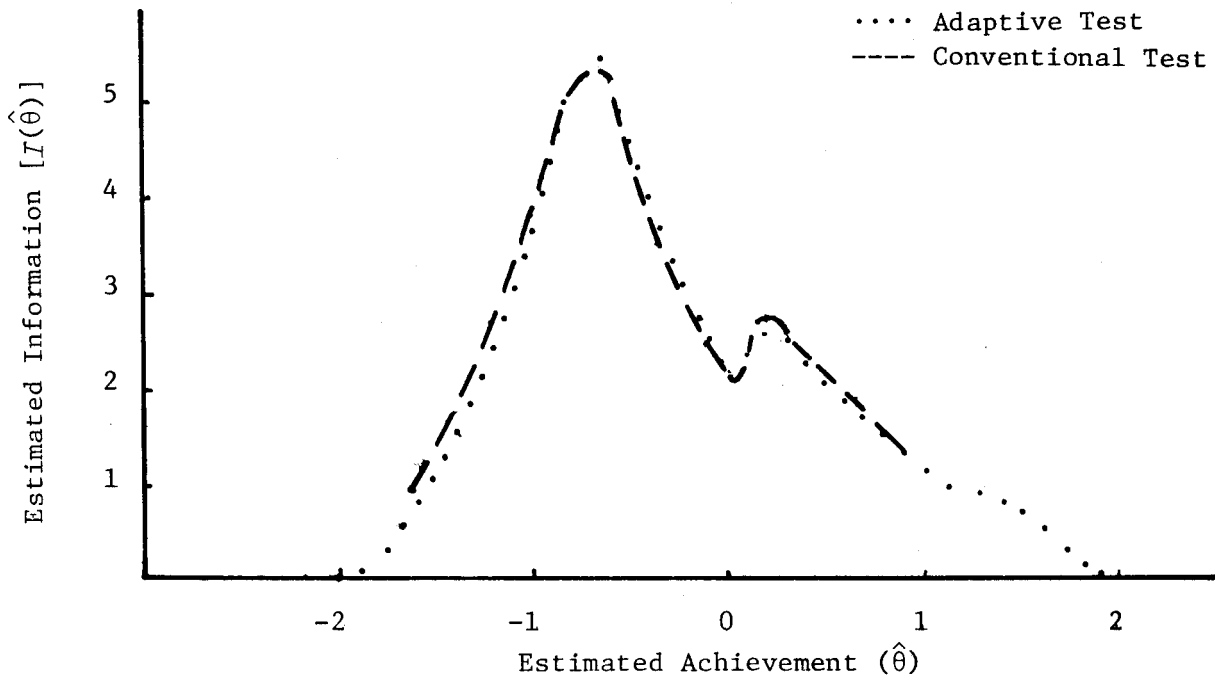


Figure C-4
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 5

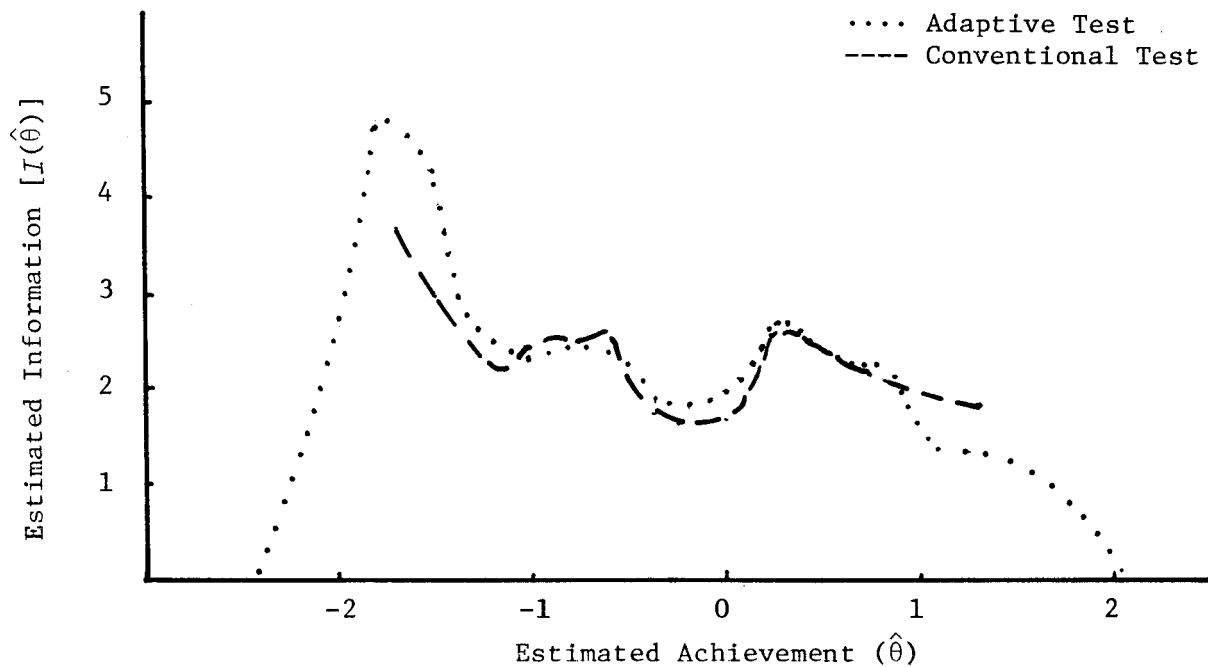


Figure C-5
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 6

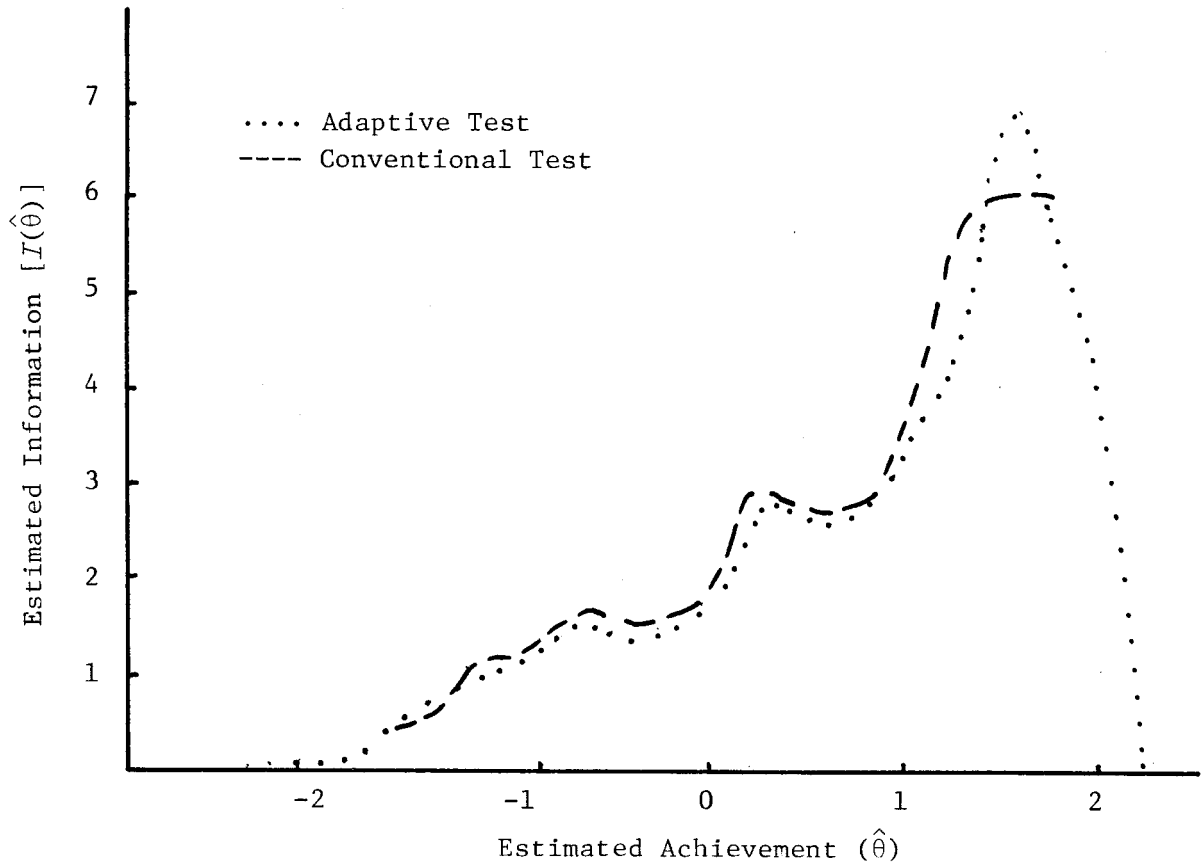


Figure C-6
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 7

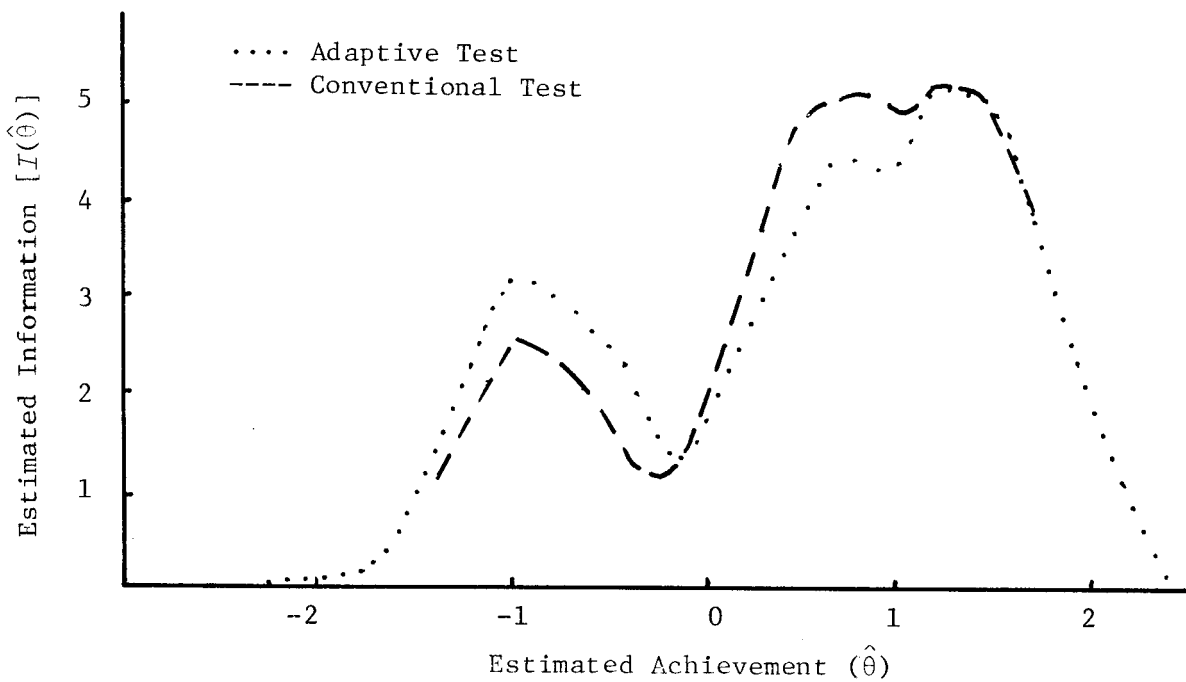


Figure C-7
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 8

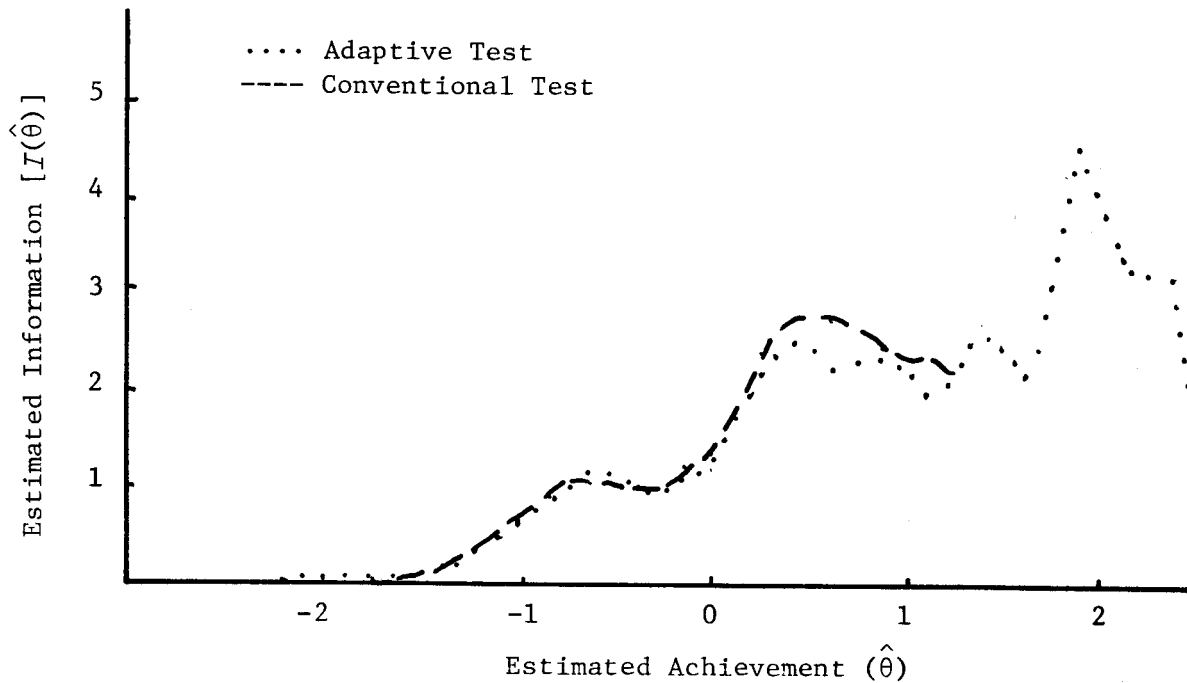


Figure C-8
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 9

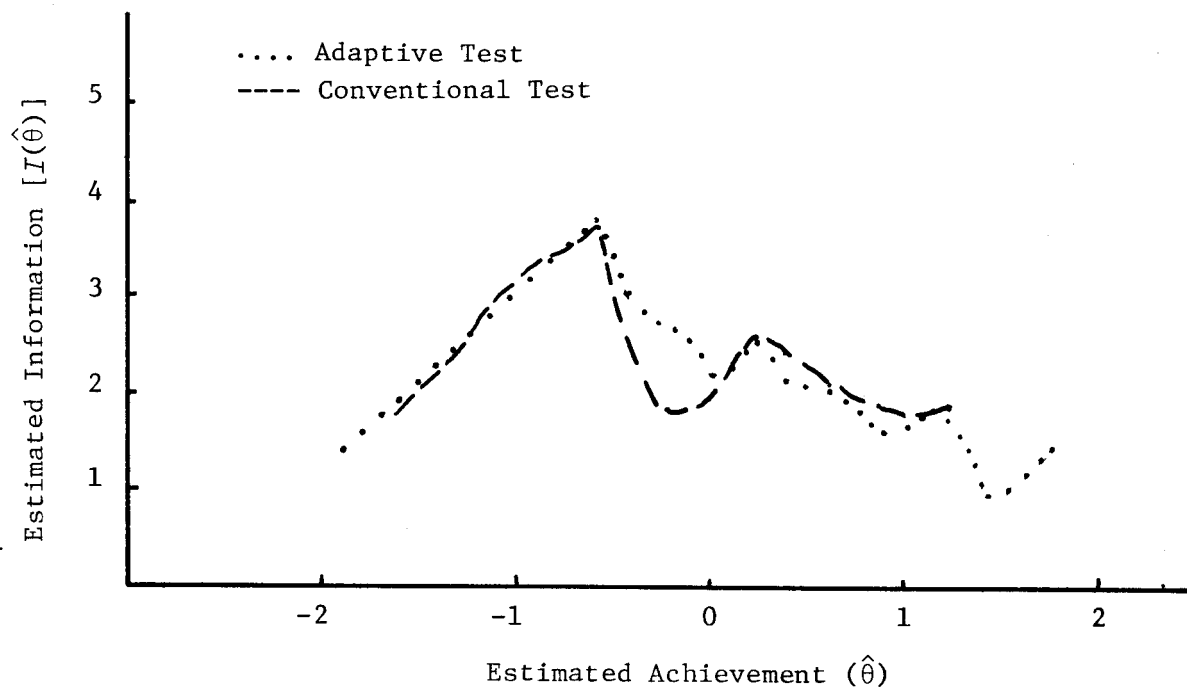


Figure C-9
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 10

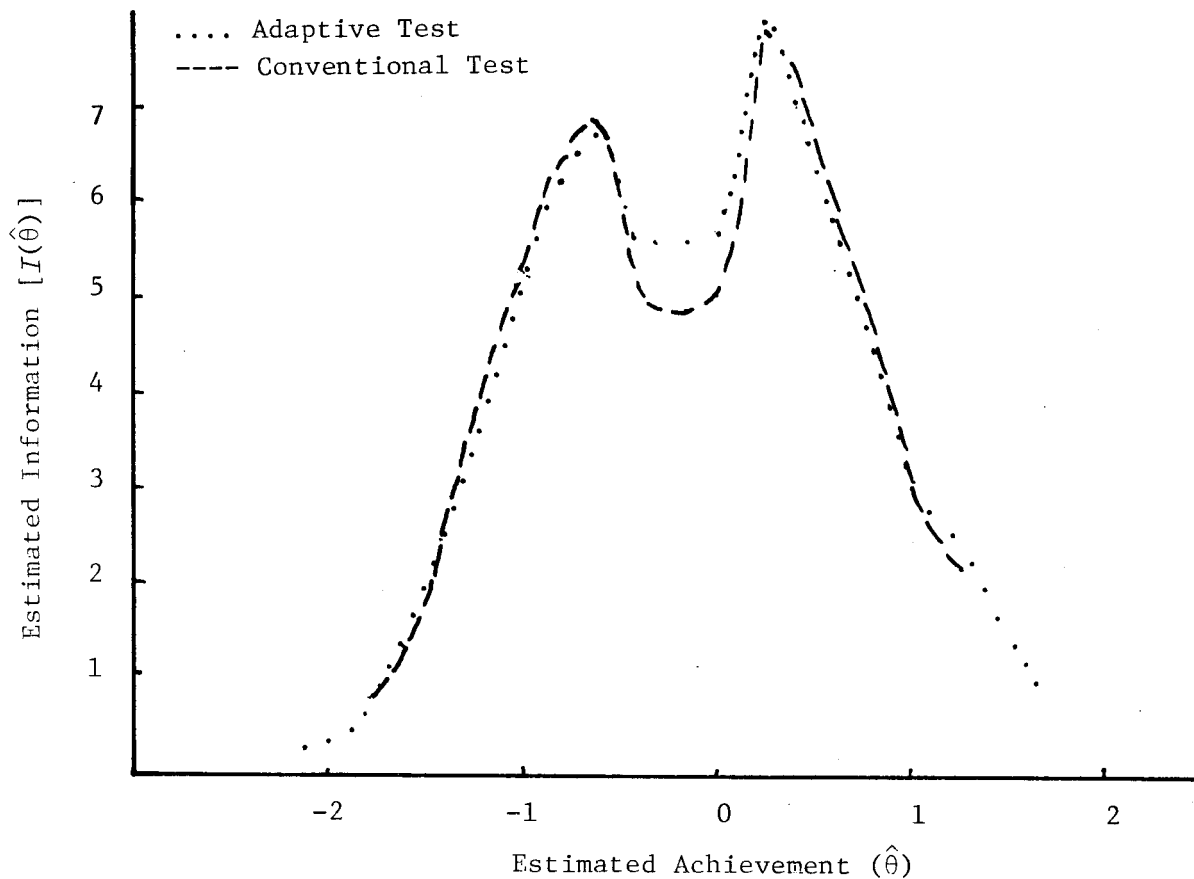


Figure C-10
Smoothed Information Curves for
Adaptive and Conventional Tests for Subtest 11

