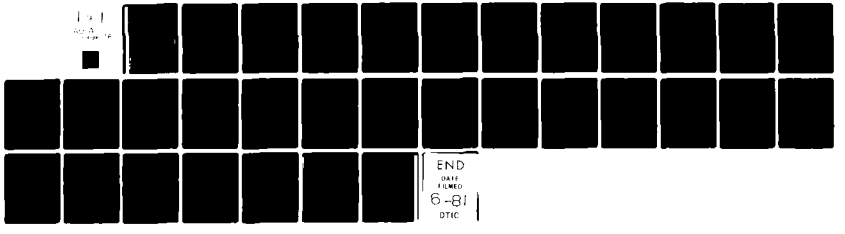
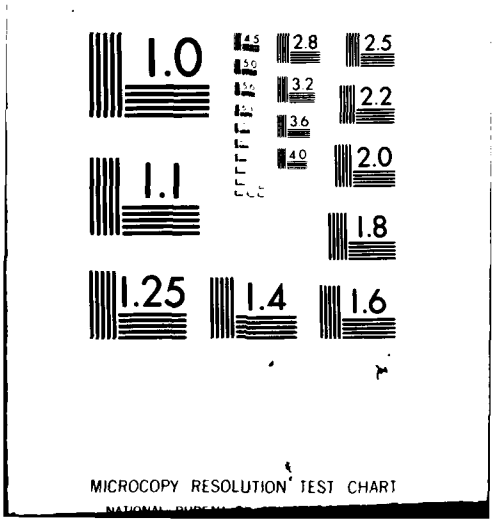
AD-A099 676 MINNESOTA UNIV MINNEAPOLIS COMPUTERIZED ADAPTIVE TES--ETC F/6 5/10
COMPUTERIZED ADAPTIVE ABILITY TESTING.(U)
APR 81 D J WEISS N00014-79-C-0324

UNCLASSIFIED NL



END
DATE FILMED
6-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

AD A099676

Final Report: COMPUTERIZED ADAPTIVE ABILITY TESTING

David J. Weiss

April 1981

DTIC
ELECTE
JUN 3 1981
S D
A

COMPUTERIZED ADAPTIVE TESTING LABORATORY
PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS MN 55455

Final Report of Project NR150-431, N00014-79-C-0324
Supported by funds from the
Office of Naval Research, the Air Force Human Resources Laboratory, Air Force Office of Scientific Research, Army Research Institute, and monitored by the Office of Naval Research
David J. Weiss, Principal Investigator

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

DTIC FILE COPY

81 6 03 021

Unclassified

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

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A099 676	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Final Report: Computerized Adaptive Ability Testing		5. TYPE OF REPORT & PERIOD COVERED Final Report. 1 Apr 1979 to 30 Jun 1980
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0324
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE: 61153N Proj: RR042-04 TA: RR042-04-01 WU: NR150-431
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE April 1981
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassifed
		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 Office of Naval Research, the Air Force Human Resources Laboratory, Air Force Office of Scientific Research, and the Army Research Institute, and monitored by the Office of Naval Research.

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

Testing	Sequential Testing	Programmed Testing
Ability Testing	Branched Testing	Response-Contingent Testing
Computerized Testing	Individualized Testing	Automated Testing
Adaptive Testing	Tailored Testing	Item Characteristic Curve Theory
Item Response Theory	Latent-Trait Test Theory	Person Fit

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

The objectives and approach of this 15-month research program are described. These objectives included (1) evaluation of the performance of adaptive testing strategies under conditions that more reasonably represent the conditions under which the strategies might be applied in live testing, including effects of errors in item parameters on the performance of adaptive testing strategies and comparisons of adaptive testing strategies in live testing; (2) evaluation of the utility for adaptive testing of a number of test item formats and response modes that might be used as replacements for the multiple-choice item;

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

Unclassified

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

411701

(3) investigation of the utility of a number of person-fit indices designed to identify lack of fit of individuals to item response theory models; and (4) investigation of the potential of several cognitive information-processing types of tasks for computerized adaptive administration. Research approaches and preliminary results are summarized for each objective. Additional research plans currently being implemented under Project NR 150-433 with which this project was combined, are described.

CONTENTS

Introduction..... 1

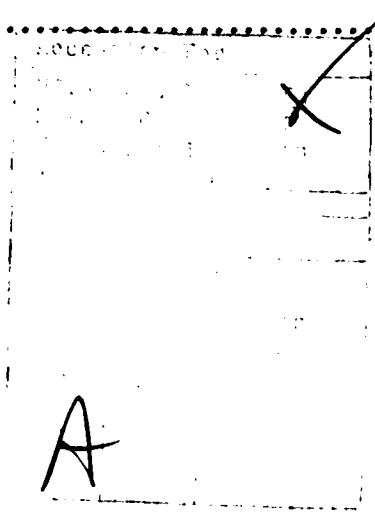
Adaptive Testing Strategies..... 1
Objective..... 1
Effects of Errors in Item Parameter Estimates on Adaptive Testing Strategies..... 2
Approach..... 2
Results..... 2
Additional Research in Progress..... 6
Live-Testing Comparison of Adaptive Testing Strategies..... 6
Approach..... 6
Results..... 7
Additional Research in Progress..... 8
Future Research Plans..... 8

Item Formats and Response Modes..... 9
Objective..... 9
Approach..... 9
Results..... 11
Additional Research in Progress..... 13

Fit of Individuals to Item Response Theory Models..... 13
Objective..... 14
Approach..... 14
Additional Research Plans..... 15

New Type of Ability Tests..... 16
Objective..... 16
Approach..... 17
Memory for Patterns..... 17
Digit Span..... 19
Analysis and Future Plans..... 20

References..... 21



FINAL REPORT:
COMPUTERIZED ADAPTIVE ABILITY TESTING

This research program was designed to study four areas relevant to adaptive ability testing:

1. Evaluation of adaptive testing branching strategies.
2. Use of different item formats and response modes.
3. Fit of individuals to item response theory models.
4. New types of ability tests designed specifically for computerized adaptive administration.

Research in pursuance of these objectives, originally scheduled for a three-year period, began on April 1, 1979, and continued through June 30, 1980, at which time the research objectives of this project were combined with those of Project NR 150-433, "Computerized Adaptive Achievement Testing."

This report summarizes the progress made during this 15-month period. No technical reports were completed during this period; technical reports begun in this project will be completed under Project NR 150-433. For each of the four objectives listed above, this report (1) describes the objective, (2) details the approaches used to study the objective, (3) summarizes results that were available at the completion of the reporting period, and (4) describes tentative plans for further research on the objective to be continued in Project NR 150-433.

Adaptive Testing Strategies

Previous simulation studies using adaptive testing have used relatively unrealistic item pools (e.g., Gorman, 1980; McBride, 1976; Reckase, 1976; Urry, 1970; Urry, 1971; Vale, 1975). These item pools have been unrealistic because they assumed that the item parameters describing the items in the pool were completely error free, as well as assuming item difficulty and distribution characteristics that did not reflect those of real ability tests. Previous studies have also been unrealistic in that they have assumed that the responses of the hypothetical testees to these items have conformed precisely to the one-dimensional latent trait model. When the results of previous simulation studies are extrapolated to real item pools constructed from real item parameters, they may not generalize, because real item pools are constructed from item parameters that include estimation error and may deviate substantially from unidimensionality.

Objective

The objectives of this research program were to evaluate the performance of adaptive testing strategies under conditions that more reasonably represent the conditions under which these strategies might occur in live-testing applications and to compare findings from selected simulation studies to those obtained in live testing. Research during the reporting period was concerned with (1) ef-

fects of errors in item parameters on the performance of adaptive testing strategies and (2) live-testing comparisons of adaptive testing strategies.

Effects of Errors in Item Parameter Estimates on Adaptive Testing Strategies

Approach. This objective was pursued by means of monte carlo simulation studies that built upon empirical information regarding the nature and extent of errors in item parameter estimates due to the numbers of testees and items on which the item response theory (IRT) item parameters were estimated. Data on the kinds and degrees of error associated with IRT item parameterization techniques by different item parameterization methods were modeled in the monte carlo simulations. The kinds and degrees of errors observed in real data item parameterization were translated into the monte carlo simulation model and served as independent variables in a series of studies systematically varying the magnitude and kind of item parameterization estimation error for the difficulty, discrimination, and "guessing" parameters separately and in combination. Dependent variables in these studies were test information, bias, correlation of ability estimates with true ability, and other characteristics of the ability estimates derived from the application of selected adaptive testing strategies; two conventional tests were also included in the study for comparison purposes. The studies were also designed to use an item pool that realistically reflected the composition of real item pools used in actual ability tests, in terms of the distributions of the IRT item parameter estimates.

Figure 1 summarizes the design of this study. Using a three-parameter IRT model and an item pool designed to reflect an adaptive testing item pool that had been used in a live-testing study, monte carlo data were generated for 100 simulees at each of 17 levels of ability, ranging from $\theta = -3.2$ to $\theta = +3.2$. Based on data available in this IRT item parameterization literature, varying degrees of error were added to the parameter estimates for item discrimination (a), difficulty (b), and "guessing" (c). Table 1 shows the item parameter sets used in this study: Set 1 was the baseline comparison data set in which there was no error in the item parameter estimates; in Sets 11, 12, and 13 varying amounts of error were added to the a parameter; in Sets 21, 22, 23, and 24 error was added to the b parameter; Sets 31 and 32 added errors to the c parameter; and in Sets 41 and 42 errors occurred in all three parameters simultaneously.

Using the error-laden item parameter sets, three types of adaptive tests (stratified adaptive, or stradaptive; maximum information; and Bayesian) were administered to each of the 1,700 simulees. All tests were scored by maximum likelihood at test lengths of 5 to 30 items, in increments of 5 items. In addition, both peaked and rectangular conventional tests were constructed using classical test construction procedures; and these tests, along with the adaptive tests using the error-free item pool, were also administered to the 1,700 simulees. Testing strategies were compared in terms of fidelity (the correlation of true and estimated θ levels), observed and theoretical information, efficiency, inaccuracy, bias, and root mean square error (RMSE) for the θ estimates.

Results. Table 2 presents a selection of the results for four of the dependent variables. The fidelity measure was computed on a normally distributed sample of 300 simulees; data for the other criterion measures were averaged across the 1,700 simulees. As Table 2 shows, with the exception of Item Set 42,

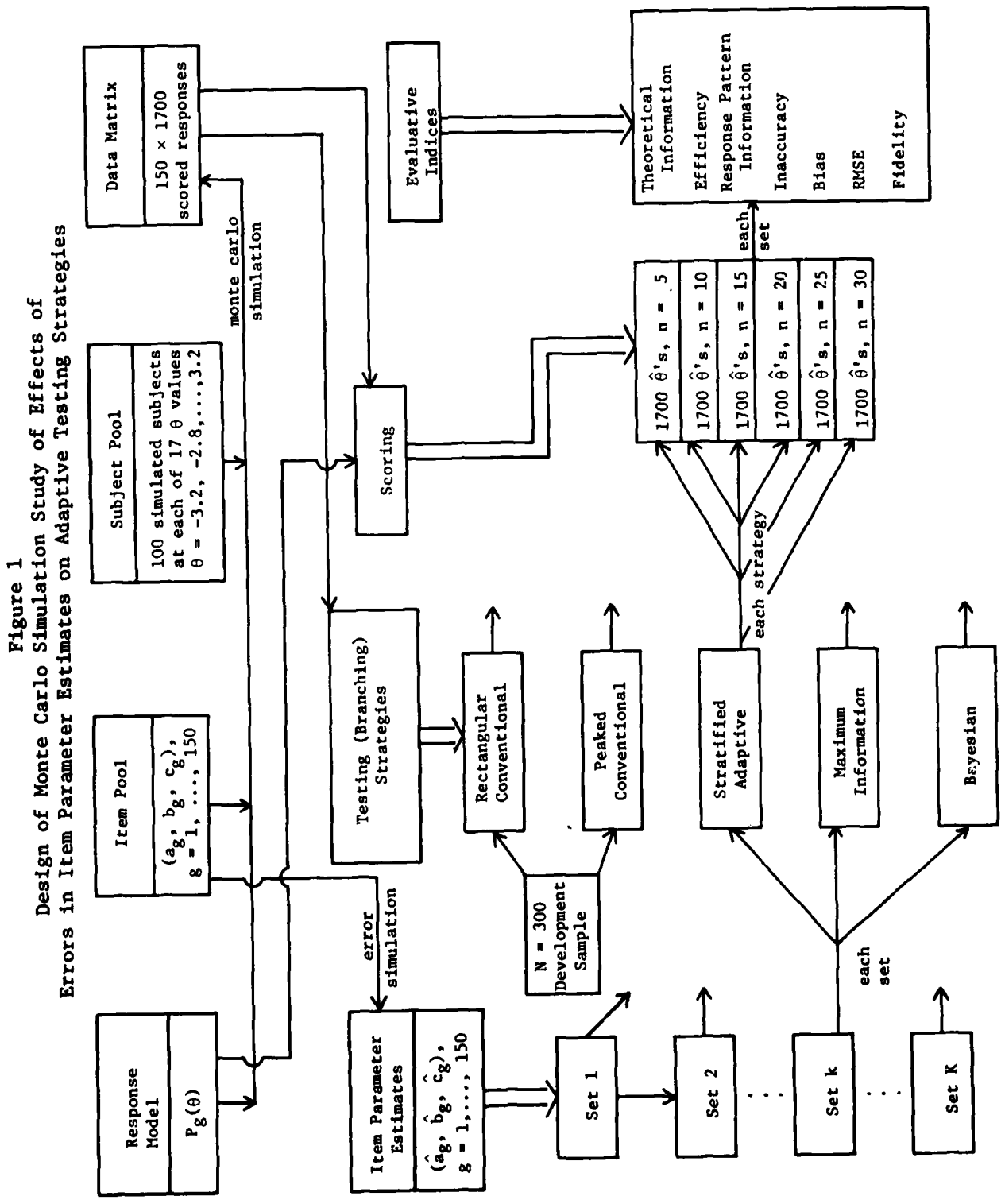


Figure 1
Design of Monte Carlo Simulation Study of Effects of
Errors in Item Parameter Estimates on Adaptive Testing Strategies

Table 1
Error Simulated in Item Parameter Estimate Sets

Item Set	Description	Specified RMSE a	Specified RMSE b	Specified RMSE c	Obtained RMSE a	Obtained RMSE b	Obtained RMSE c	Obtained $r(p,\hat{p})$ a	Obtained $r(p,\hat{p})$ b	Obtained $r(p,\hat{p})$ c
1	Error-Free Item Set	.00	.00	.00	.00	.00	.00	1.00	1.00	1.00
11	Small Error in a	.20	.00	.00	.22	.00	.00	.75	1.00	1.00
12	Moderate Error in a	.40	.00	.00	.39	.00	.00	.56	1.00	1.00
13	Large Error in a	.60	.00	.00	.52	.00	.00	.37	1.00	1.00
21	Moderate Error in b	.00	.10	.00	.00	.09	.00	1.00	.99	1.00
22	Large Error in b	.00	.30	.00	.00	.30	.00	1.00	.98	1.00
23	Extreme Error in b	.00	1.00	.00	.00	.89	.00	1.00	.88	1.00
24	Very Large Error in b	.00	.50	.00	.00	.48	.00	1.00	.96	1.00
31	Moderate Error in c	.00	.00	.04	.00	.00	.04	1.00	1.00	.70
32	Large Error in c	.00	.00	.08	.00	.00	.08	1.00	1.00	.46
41	Worst Probable Combined Error	.60	.30	.08	.51	.32	.08	.47	.98	.46
42	Extreme Combined Error	.60	.00	.08	.58	.97	.08	.44	.88	.43

which represented extreme (and probably unrealistic) levels of error in all three item parameters, the adaptive tests with error-laden item parameters achieved higher fidelities at all test lengths than did the peaked (P) and rectangular (R) conventional tests, with larger differences occurring for shorter test lengths. There were virtually no differences in fidelities for the adaptive strategies at 20- or 30-item test lengths, with a tendency for the maximum information (MI) adaptive test to perform somewhat more poorly than the stratified adaptive (SA) or Bayesian (B) tests at 10-item test lengths. Results for the other dependent measures tended to support the fidelity analysis; that is, with the exception of Item Set 42, adaptive tests using error-laden item parameter estimates generally achieved scores with lower levels of inaccuracy, bias, and RMSE than did conventional tests of the same lengths using error-free item estimates.

Analyses of the data in terms of dependent measures conditioned on values of θ--inaccuracy, bias, RMSE, the two information measures, and efficiency--supported the findings from the overall analysis. When errors occurred in the a, b, and c parameters separately, there was very little effect on these indices and the adaptive tests measured better than the conventional tests at virtually all levels of θ. There was essentially no measurement degradation as the result of errors in c and a, with a slightly greater effect for b. For realistic val-

Table 2

Overall Measures at 10-, 20-, and 30-Item Test Lengths for Varying Amounts of Error in a, b, and c, and All Three Simultaneously, for Rectangular (R) Conventional and Peaked Conventional (P) tests, and for Stradaptive (S), Maximum Information (MI), and Bayesian (B) Adaptive Tests

Measure and Number of Items	R	P	Item Set 13: RMSE a = .52, RMSE b = .00, RMSE c = .00			Item Set 22: RMSE a = .00, RMSE b = .30, RMSE c = .00			Item Set 23: RMSE a = .00, RMSE b = .89, RMSE c = .00			Item Set 32: RMSE a = .00, RMSE b = .00, RMSE c = .08			Item Set 41: RMSE a = .51, RMSE b = .32, RMSE c = .08			Item Set 42: RMSE a = .58, RMSE b = .97, RMSE c = .08		
			SA	MI	B	SA	MI	B	SA	MI	B	SA	MI	B	SA	MI	B	SA	MI	B
Fidelity																				
10	.83	.73	.88	.84	.90	.91	.90	.93	.86	.86	.87	.93	.90	.93	.86	.85	.89	.82	.78	.81
20	.92	.89	.94	.95	.95	.96	.96	.96	.92	.93	.93	.96	.96	.96	.94	.95	.95	.89	.89	.89
30	.94	.93	.96	.97	.97	.97	.97	.97	.95	.95	.95	.97	.97	.97	.96	.97	.96	.93	.92	.93
Inaccuracy																				
10	.53	.70	.40	.44	.40	.40	.42	.36	.51	.65	.48	.35	.38	.37	.46	.47	.46	.70	.78	.56
20	.38	.48	.29	.28	.27	.29	.27	.26	.41	.46	.39	.26	.24	.28	.31	.30	.32	.62	.61	.46
30	.33	.41	.24	.24	.24	.25	.23	.23	.38	.42	.38	.21	.20	.22	.26	.25	.25	.55	.54	.43
Bias																				
10	-.08	-.18	-.08	-.03	-.07	-.05	.00	-.03	.03	-.13	.04	-.06	-.12	-.13	-.04	-.15	-.14	-.15	-.25	-.16
20	-.04	-.15	.00	.04	-.01	-.02	.01	-.03	.12	.10	.09	-.02	-.04	-.06	-.02	-.03	-.08	-.16	-.11	-.16
30	.02	-.10	.04	.06	.02	.00	.02	.00	.12	.13	.11	-.01	-.02	-.03	.00	.02	-.03	-.12	.01	-.10
RMSE																				
10	.67	.98	.55	.64	.56	.55	.58	.47	.64	.82	.62	.48	.57	.53	.64	.67	.66	.87	.98	.71
20	.48	.69	.38	.37	.35	.37	.36	.35	.51	.57	.49	.33	.32	.35	.41	.40	.44	.75	.74	.59
30	.44	.59	.32	.31	.30	.33	.30	.31	.48	.52	.48	.28	.27	.29	.34	.32	.34	.66	.64	.53

ues of combined error in the three item parameters, some measurement degradation occurred for the adaptive tests, but they still measured better than the rectangular conventional test at all θ levels, and better than the peaked conventional test for about three-fourths of the θ scale. There were few consistent differences in the performance of the different adaptive testing strategies.

Additional research in progress. Results of this study indicated very little effect of errors in item parameter estimates on the measurement performance of adaptive testing strategies. Since this study was the first to investigate this question, it was necessarily limited in a number of ways. Consequently, further simulations are planned that (1) vary the characteristics of the item pool used in order to determine the generality of the findings across item pools with different characteristics, in terms of levels of the three item parameters; that (2) allow correlated errors to occur in the item parameter estimates, since only uncorrelated errors were used in this study; and that (3) examine the effects of error in item parameter estimates separately for one-, two-, and three-parameter IRT models.

Live-Testing Comparison of Adaptive Testing Strategies

Approach. Three testing strategies--peaked conventional, Bayesian adaptive, and maximum information adaptive--were compared on the basis of alternate forms reliability and observed information. The tests were composed of 60 five-choice vocabulary items that were divided into two 30-item alternate forms. The conventional test was peaked in information values evaluated at $\theta = 0.0$. Items administered in the maximum information and Bayesian testing strategies were selected according to their adaptive item selection routines. There were 373 students in the conventional testing condition, 390 in the Bayesian testing condition, and 233 in the maximum information testing condition.

Testing strategy was the major independent variable of interest. Methods of scoring were also compared. These included logistic maximum likelihood scoring, Bayesian scoring, and (for the conventional test) proportion-correct scoring. Test length was a third independent variable of interest. Thirty test lengths were obtained by scoring each 30-item test at each test length from 1 to 30 items. Testing strategies were compared on the basis of alternate forms reliability by correlating corresponding ability estimates obtained from Forms A and B for a given testing strategy.

Since the test data were scored in at least two ways (Bayesian and maximum likelihood), a total of seven combinations of testing strategy and scoring method were compared on the basis of alternate forms reliability. Scoring strategy was compared on the basis of alternate forms reliability by comparing reliabilities of a single testing strategy scored by more than one method. Three of the alternate forms reliabilities paired the appropriate scoring method with each of the three testing strategies. These were proportion-correct scoring of conventional tests, maximum likelihood scoring of maximum information tests, and Bayesian scoring of Bayesian-administered tests. The remaining four alternate forms reliabilities were obtained by scoring the item response data by a scoring routine other than the appropriate one. In this way, reliabilities were obtained for the Bayesian scoring of the maximum information test, maximum likelihood scoring of the Bayesian test, Bayesian scoring of the conventional test, and maximum likelihood scoring of the conventional test. Reliabilities were

calculated as a function of test length. Scoring method correlations were obtained by correlating estimates obtained from different scorings of the same testing strategy. These correlations were used to analyze the similarity of ability estimates obtained from different scoring methods applied to a single set of data.

The three testing strategies were also compared on the basis of their errors of measurement. This was assessed in two ways: (1) using estimated errors of measurement derived from maximum likelihood scoring and (2) using estimated errors of measurement from Bayesian scoring. In the first method, test item responses were scored by maximum likelihood methods, and the standard errors of measurement (SEM) associated with each ability estimate was calculated. These values are the reciprocal of the square root of test information at a given $\hat{\theta}$ level and estimate the standard deviation of the estimated θ values around the true θ value; the larger the SEM, the more likely the estimate will be inaccurate. The posterior variance of the Bayesian ability estimate was the second index used to compare the testing strategies on the basis of measurement accuracy. Both the SEM and the posterior variances were examined as a function of estimated ability level.

Results. A preliminary report of the results of this study is in Johnson and Weiss (1980). Parallel forms reliabilities of the three testing strategies showed that after 11 items the peaked conventional test yielded higher reliabilities than either of the adaptive tests. The greatest difference between reliabilities was $\underline{r} = .09$ between the adaptive and conventional tests at the 30-item test length; the reliabilities of the adaptive tests were $\underline{r} = .81$, compared with the final reliability of $\underline{r} = .90$ for the conventional test. The conventional test reliability was nearly identical to that of the Bayesian test up to the 10-item test length, but after that point the conventional test reliability increased more quickly than that of the adaptive tests. Although adaptive test reliabilities showed signs of leveling off toward the end of the test, the reliability of the conventional test appeared to increase steadily.

In comparisons of testing strategies scored by other than optimal scoring strategies, the Bayesian scoring of the conventional and maximum information testing strategies yielded higher reliabilities than the maximum likelihood scoring of the conventional and Bayesian testing strategies. These data indicate that Bayesian scoring of an adaptive test may yield more stable estimates of ability than maximum likelihood scoring. The data also illustrate the inappropriateness of scoring conventional tests with maximum likelihood scoring methods, since extremely low reliabilities (maximum of $\underline{r} = .75$) were obtained at all test lengths. The correlations between scores on the same testing strategy scored by different methods showed that the highest correlations were obtained for Bayesian and proportion-correct scores of the conventional test, with most correlations between .97 and .99. The second highest level of correlation was between the Bayesian and maximum-likelihood-scored maximum information test, with most correlations between .93 and .95. When the maximum information adaptive test was scored by the Bayesian scoring method, reliabilities of short adaptive tests were higher than those of the conventional test, and differences in reliabilities were smaller at longer test lengths.

On the basis of the reliability data, few conclusions can be drawn about the relative merits of the three testing strategies. Limitations of the item

pool might account in part for the lowered reliability of the adaptive tests in comparison to the conventional test, since adaptive tests depend heavily on the quality of the items in the item pool. The item pool used for the two adaptive tests had fewer items at the extremes of the ability range, and these items had relatively lower discrimination parameters. Especially at abilities where there were fewer items, it is likely that the correlations between ability estimates would be attenuated and that the adaptive process would be at a disadvantage as testing progressed. The result would be that toward the end of testing there would be fewer and fewer items available at a given ability level.

Another factor that limits the comparison of the testing strategies in terms of alternate forms reliability correlations is the distribution of ability in the population. Since values of the Pearson product-moment correlations depend on the distributions of the ability estimates involved, different ability distributions can result in different levels of correlation. Thus, the reliability correlations confound the distribution of the ability estimates with the measurement precision of the testing strategies.

Errors of measurement derived from test information yield comparisons of testing strategies that are unconfounded by the distribution of the ability estimates. Comparisons of the testing strategies on the basis of SEMs and posterior variances showed that at no point on the ability continuum were the errors of measurement smaller in the conventional test than in the adaptive tests. In both error of measurement comparisons there was poorer measurement at the low end of the ability distribution, although the extremes--both positive and negative--were less precisely measured than the center of the ability continuum. The results indicate that the adaptive tests yielded about the same level of measurement precision and that these levels were greater than those obtained from the conventional test at all levels of ability. Thus, adaptive testing strategies yielded scores with greater precision/information (lower errors of measurement) than did the conventional testing strategy.

Additional research in progress. Since the reliability results of this study were contrary to expectations and conflicted with other research using a similar design but different tests (Kingsbury & Weiss, 1980; McBride, 1980), a fourth test was added to the study. To examine the effects of test difficulty on the results, this test was a second conventional test in which average item difficulty was higher than that of the first conventional test. Data were collected using this test from 530 students on a 60-item conventional test consisting of two embedded 30-item alternate forms. The alternate forms reliabilities of these tests will be computed at test lengths from 1 to 30 items, and the data will be further analyzed to permit direct comparisons with the three other testing strategies.

Future Research Plans

In addition to using item parameters that contain varying degrees of error, real adaptive testing item pools may deviate from the unidimensional IRT model that has been applied in all adaptive testing simulations. Since deviations from unidimensionality (e.g., Bejar, 1979, 1980; Bejar, Weiss, & Kingsbury, 1977; Reckase, 1978) can potentially affect the performance of adaptive testing strategies, a series of monte carlo simulation studies will be constructed around the degrees and types of dimensionality observed in ability test data.

These studies will consist of the generation of testee responses using the underlying multidimensional structures observed in ability test items, but adaptive branching will occur by means of several adaptive testing strategies designed for unidimensional adaptive testing. Thus, the research question will be the effects of violation of the unidimensionality assumption on the performance of adaptive testing strategies. Again, the evaluative criteria will consist of information, bias, correlation of true ability and ability estimates, and other characteristics of the ability estimates derived from the unidimensional adaptive testing strategies.

Item Formats and Response Modes

The use of interactive computers to administer ability tests allows the design and use of test items that do not make use of the typical multiple-choice item format. Research on alternatives to the typical multiple-choice item (Bejar, 1975; Vale, 1977) suggests that there is considerable improvement possible in information utilization from adaptive testing by use of response modes other than multiple-choice items. Consequently, continued research in this area was indicated.

Objective

To evaluate the utility for adaptive testing of a number of response modes and item formats usable in adaptive testing.

Approach

This objective was pursued using the six item types shown in Figure 2. The studies were concerned with the following characteristics of these item types:

1. The relationship of responding in the various formats to ability levels.
2. The reliability of test item responses and ability estimates obtained in the various formats.
3. Information characteristics of items and tests utilizing the various formats.
4. The relationships among test item responses using the various formats.
5. The relative validity of responses obtained from the various formats.
6. The generality of findings obtained from the different response formats to different populations and different ability dimensions.
7. The comparative factor structure of tests administered in the various formats.

The research was designed to study the characteristics of the six item formats in several ability areas. It is essentially a search for an item format that allows testees to express as much knowledge as they have available about a given question while minimizing the effects of guessing. The results obtained from this series of studies will be used to select several item formats to be used in computerized adaptive testing.

Two sets of 30 multiple-choice items were chosen from available item pools. One set of items was chosen from a pool of analogy items, and the second set was chosen from a pool of arithmetic reasoning items. Both item pools included item

Figure 2
Description of Response Formats

1. Multiple-choice items with conventional response format. These items were conventional multiple-choice items with four alternatives. The examinee was asked to choose the correct answer.

   Example. Procedure : Activity
   1. Diplomacy : Tact
   *2. Itinerary : Journey
   3. Minutes : Committee
   4. Index : Book

2. Multiple-choice items with probabilistic response format. These items were exactly the same as the conventional multiple-choice items, but the examinees were asked to assign 100 points among the four alternatives to indicate their confidence in the "correctness" of each alternative.

Example. Procedure : Activity	(Possible Answer)
1. Diplomacy : Tact	0
2. Itinerary : Journey	75
3. Minutes : Committee	25
4. Index : Book	0

3. Dichotomous items with a yes-no (dichotomous) response format. For these items only the item stem and one alternative were presented. The examinees were asked to respond with a "yes" if they thought the alternative provided was a correct answer to the question, and "no" if they thought it was not.

Example. Q. Procedure : Activity	(Possible Answer)
A. Index : Book	Yes

4. Dichotomous items with a probabilistic response format. These items were identical to the dichotomous items with a yes-no response format, but, the examinees were asked to respond with a probability (a number from 0 to 100) which reflected their confidence that the alternative provided was a correct answer to the question.

Example. Q. Procedure : Activity	(Possible answer)
A. Index : Book	10

5. Free-response items with a conventional response format. For these items only the item stem was presented. The examinees were asked to provide their own answers.

   Example. Q. Procedure : Activity
   Itinerary : _____

6. Free-response items with a probabilistic response format. Once again only the item stem was presented, but this time the examinees were asked to provide an answer to the question and to assign a probability (a number from 0 to 100) to the answer they gave to indicate their confidence in the "correctness" of their response.

Example. Q. Procedure : Activity	(Possible answer)
Itinerary : _____	Trip 90

parameters calculated on large numbers of individuals using the three-parameter logistic model of LOGIST (Wood & Lord, 1976; Wood, Wingersky, & Lord, 1976). The items were chosen to represent a uniform range of difficulty and discrimination parameters. Each set of 30 items was then modified to conform to the item formats shown in Figure 2. For each item, the item stem remained the same, while the response formats were changed, resulting in the 6 sets of 30 items, each with a different response format. Tests were then constructed utilizing these items and the tests were administered in various combinations to a number of groups of several hundred college students.

The data collected will be used to determine the factor structure and the convergent and discriminant validity of the tests using the different response formats in the two ability domains and will be compared with similar data already available on vocabulary ability. The data will also be used to compare the item parameters, item and test information functions, internal consistency, and interrelationships among scores derived from the various response formats.

Results

Preliminary data analyses were completed on data collected on three of the six item formats using the analogies items. The formats analyzed were multiple-choice items with conventional response format (MCC), multiple-choice items with probabilistic response format (MCP), and dichotomous items with a dichotomous (MCD) response format (Types 1 through 3 in Figure 2).

Examination of Table 3 shows the average item scores for the 30 items in the three formats. The MCC and MCP items were rank ordered very similarly, as evidenced by the correlation of .91 between the average item scores for these two formats. The DD items ordered themselves somewhat differently, as indicated by the correlations between the DD average item scores and the MCC and MCP average item scores, which were .43 and .52, respectively. Although the present results reflect only one scoring system for the MCP items, several other scoring methods will also be investigated for this response format.

Table 4 shows validity and internal consistency reliability coefficients obtained for the three item formats. The validity coefficient reported is the correlation of total score with the reported grade-point average of the students (later analyses are planned using actual GPA rather than reported GPA). The validity coefficients for all three formats were not significantly different from each other, and were significantly different from zero. The MCP items had the highest internal consistency reliability, but the lowest validity. The MCC items had the second highest reliability, and highest validity; and the DD items had lowest reliability with moderate validity.

A number of factor analyses were performed on the data from the three item formats. Both principal axes and confirmatory analyses were used. The principal axes analyses showed that two orthogonal factors were extracted for each response format, and the pattern of positive and negative loadings was extremely similar for the MCC and MCP items but was very different for the DD items. In addition, the percent of total score variation accounted for by the two-factor solution varied with item format.

Confirmatory factor analysis was performed using the principal axes factor

Table 3
Average Item Scores for the Same 30 Analogies Items in Multiple-Choice Conventional Format (MCC), Multiple-Choice Probabilistic Format (MCP), and Dichotomous-Dichotomous (DD) Format

Item Number	Response Format		
	MCC*	MCP**	DD*
1	.86	1.18	.87
2	.64	.66	.42
3	.78	.99	.65
4	.52	.58	.81
5	.73	1.02	.80
6	.79	1.08	.52
7	.28	.24	.64
8	.48	.64	.61
9	.61	.92	.51
10	.58	.80	.73
11	.60	.73	.44
12	.89	1.19	.71
13	.77	1.08	.79
14	.62	.77	.32
15	.75	1.29	.82
16	.78	1.07	.92
17	.64	.94	.70
18	.84	1.48	.87
19	.41	.63	.80
20	.68	1.05	.84
21	.87	1.14	.98
22	.84	1.06	.65
23	.87	1.32	.78
24	.91	1.41	.98
25	.68	1.08	.56
26	.77	.88	.80
27	.75	1.25	.83
28	.86	1.29	.94
29	.54	.60	.58
30	.35	.42	.66

*Proportion correct.
**Average score with range from 2.00 to -1.00.

solution for the MCC items as the basis for the model. The data showed better fit to the model for the MCP items than for the DD items, with the second factor a rather inconsequential factor for all three response formats.

Thus, the results obtained thus far advise against the use of the DD item format. This item format is less reliable than the other response formats, has only moderate validity, shows high levels of guessing, and does not appear to be consistently measuring analogies ability. The DD response format does not appear to be a viable alternative to the multiple-choice item. On the other hand, the MCP item format does appear to be a promising alternative to the traditional

Table 4
Alpha Internal Consistency Reliability Coefficients and Validity Correlations with Reported GPA for Three Response Formats

Response Format	N	Coefficient Alpha	Validity Coefficient
Multiple-Choice Conventional	486	.85	.23
Multiple-Choice Probabilistic	299	.91	.17
Dichotomous-Dichotomous	303	.59	.20

multiple-choice item. It is more reliable, nearly equally as valid, and appears to be measuring analogies ability to a greater degree than the conventional multiple choice-items.

Additional Research in Progress

Further research and analysis, of course, are needed to determine whether or not these results are generalizable across ability domains, and whether any of the item formats not yet analyzed will also show promise as alternatives to the standard multiple-choice item. Considerable amounts of additional data will be obtained on these item formats using both the analogies and numerical reasoning items, and the resulting data will be analyzed to investigate the questions raised earlier. In addition, further evidence of the generality of the findings will be sought using vocabulary ability items that are being analyzed by similar methodologies. Once remaining item types are identified as replacements for the multiple-choice item, adaptive tests using these item types will be designed and their characteristics investigated.

Fit of Individuals to Item Response Theory Models

Previous research on the person response curve (Trabin & Weiss, 1979) and related research on person fit (Levine & Drasgow, 1980; Levine & Rubin, 1979) promises the capability of identifying individuals who, on a given test, are not responding in accordance with a given IRT model. This lack of fit to the model can derive from a number of possible causes, including the following:

1. Lack of motivation to respond appropriately to the test;
2. Inappropriate or nonrandom guessing, or lack of guessing;
3. Responses that are not in accord with the unidimensionality assumption of ICC theory.

Knowledge that an individual is not responding according to the model for any of these reasons would be appropriate information to be used in applied situations suggesting that the scores of that person be carefully considered before decisions are made on the basis of those scores. It may also be an important moderator variable for use in prediction studies.

Objective

To further investigate the utility of the person response curve (PRC) concept and to identify the correlates of deviations of an individual from the IRT model.

Approach

To properly investigate the usefulness of the PRC approach to the problem of person fit, its characteristics were investigated within the context of other indices designed for similar purposes. Based on a thorough review of the relatively small literature on person fit (also known as "appropriateness" measurement), the following indices were identified:

1. Trabin and Weiss's (1979) chi-square index of the fit of observed and expected PRCs.

2. Reckase's (1977) mean square deviation (MSD) index averaged over items,

$$MSD_j = \sum_{i=1}^{N} (u_{ij} - p_{ij})^2 \Big/ N \qquad [1]$$

where

MSD_j = the mean squared deviation for person j,
u_{ij} = the actual response to item i by person j,
p_{ij} = the probability of a correct response as predicted by the three-parameter IRT model, and
N = the number of items in the test.

This index is a special case of the fit of the observed PRC to the expected PRC.

3. A variation of Reckase's index in which only improbable responses are scored. It is difficult to argue that those responses that are in the predicted direction should be included in a person-fit statistic. Hence, only where $(u_{ij} - p_{ij})^2$ is greater than .25 is it included in the statistic. The divisor of the statistic is still the number of items in the test.

4. The likelihood ratio index of person fit. For a given $\hat{\theta}$ value, a probability distribution for the possible response vectors can be generated. The probability of an individual's actual response vector is divided by the probability of the most likely response vector to produce the likelihood ratio.

5. Wright (1968) proposed an index of person or item fit, which is the sum of the standardized squares of the residual after fitting the model. For the Rasch model it is $e^{(p-i)}$ for an incorrect answer and $e^{(i-p)}$ for a correct answer, where $e = 2.71$, p is $\hat{\theta}$ for an individual, and i is the difficulty of the item.

6. Response pattern information, which reflects the flatness of the likelihood function.

7. The three appropriateness indices, L_o, L_n, and $\hat{\sigma}$ described by Levine and Drasgow (1980) and Levine and Rubin (1979).

8. The difference between the difficulties of the easiest item answered incorrectly and the most difficult item answered correctly.

9. Donlon and Fisher's (1968) "personal biserial correlation" and Jacobs' (1963) variation of it.

10. The posterior variance of IRT-based Bayesian ability estimates.

The utility of these indices for identifying individual nonfit to IRT models is first being studied in simulation. As a first step, the null distributions of these person-fit statistics are being examined, to serve as reference points for the later studies to determine how well each statistic identifies person nonfit. Using a set of 2,500 simulees rectangularly distributed at 25 equally spaced levels of θ, 625 items were administered to each simulee. Items were rectangularly distributed in $\underline{b}$ with 25 items at each of 25 levels of $\underline{b}$ corresponding to the 25 levels of θ. Within each level of $\underline{b}$, items varied in discrimination at .07 intervals. Simulation data were generated separately for the 2,500 simulees for $\underline{c}$ = 0.0, .20, and .25 in order to examine the effects of guessing on the null distributions. To examine the effects of test length and discrimination, shorter tests were selected from the item pool at differing levels of $\underline{a}$. For each of these test configurations, null distributions were computed for each of the person-fit indices.

After distributions of these indices are known for model-conforming data, increasing amounts of random responses will be added to the response vectors to simulate random responding, inattention or low motivation, and guessing. Changes in the distributions that result from increasingly random responding will be noted. In this phase, percent of random response is an additional independent variable. Those person-fit statistics that are affected most strongly by random responding will be retained as good candidates for further live-data research. Results of the data analysis are not yet available from these studies.

Additional Research Plans

The effect of multidimensionality on the person-fit indices will be examined. This will be studied by generating response data from a θ known to correlate to varying degrees with the original θ and by inserting these responses into the response vectors. Degree of correlation as well as number of dimensions will be manipulated. Dependent variables studied will be the ability of the person-fit indices to identify the existence of the multidimensional response patterns.

Once promising person-fit indices are identified in the simulation studies, live-testing studies will be designed in which the fit of persons to the IRT model is empirically tested on a given pool of items. Experimental studies will be designed to attempt to induce deviations from the model in groups of individuals and to observe whether the person-fit indices are sufficiently sensitive to identify those deviations when they exist. In addition, the existence of natu-

rally occurring groups in which deviation from the IRT model might occur will be studied. For example, it can be hypothesized that on certain kinds of subtests (e.g., verbal ability tests) students from a non-English speaking culture would likely show significant deviations from unidimensionality. It could also be hypothesized that students who are not "test-wise" or who have a lack of familiarity with multiple-choice tests would show specific deviations from IRT models in their test performance. In addition, the generality of person-fit variables across ability dimensions will be studied in order to determine whether deviations of fit to the model for an individual are specific to an item pool or occur across different kinds of item domains.

New Types of Ability Tests

Psychometric attempts to measure individual differences in cognitive abilities during the last 60 years have produced tests of global abilities such as general reasoning, verbal and quantitative abilities, as well as more specific ability "factors" such as speed and flexibility of closure, spatial orientation, and word fluency (French, Ekstrom, & Price, 1963). Research in adaptive testing has shown that the precision and validity of ability tests can be improved by adaptive testing procedures. At the same time, cognitive psychologists have developed several standard tasks or paradigms that have been used to study the mechanisms and structure of aspects of memory, attention, and cognition (e.g., Sperling, 1960; Sternberg, 1966). Attempts to assess quantitative differences between individuals in such information-processing abilities and to relate them to more traditional psychometric measures is a fairly recent phenomenon (Chiang & Atkinson, 1976; Day, 1977; Hunt, Frost, & Lunneborg, 1973; Hunt, Lunneborg, & Lewis, 1975; Lunneborg, 1977; Rose, 1974). Carroll (1976) has provided an additional framework for relating traditional psychometric factors to their cognitive information-processing requirements. Some of this research has suggested that individual differences in such information-processing abilities can be measured reliably by interactive computers and that these abilities may add incremental validity in predicting external job criteria (Cory, 1977; Cory, Rimland, & Bryson, 1977).

In the past, tasks of the type that have been used by cognitive psychologists to measure information-processing abilities and by psychometricians to measure perceptual and spatial factors have been administered as blocks of fixed numbers of trials or replications. As a result, little is known about employing the important parameters of these tasks for adapting the difficulty level of replications to converge upon an individual's ability level. A major emphasis of the research, therefore, involved (1) studying some of these tasks from the point of view of how computerized adaptive administration could be meaningfully achieved and (2) evaluating the measurement benefits of adaptive administration of ability tests designed to utilize the unique capabilities of computerized administration.

Objective

The objectives were to investigate the application of adaptive testing techniques to improving the measurement characteristics of several cognitive information-processing tasks (e.g., short-term memory span, capacity of visual sensory memory). Computerized administration of these ability tests will be investigated as a means of modifying task presentations over time in a way that would not be possible in paper-and-pencil testing.

Approach

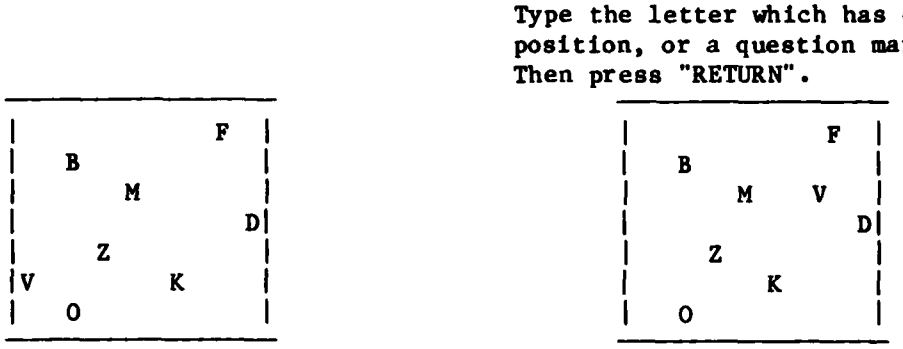
Two types of ability test items that utilized the unique capabilities of computer administration--Memory for Patterns and Digit Span--were studied to investigate the feasibility of adaptive administration of information-processing tests. These tasks were studied from the point of view of enabling the computer to adapt the difficulty of the tasks presented to the ability level of the testee during the process of testing.

Memory for Patterns. This test was based on a procedure devised by Sperling (1960) for determining the capacity of visual sensory memory. The procedure is based on presentation of arrays of letters that must be studied by the testee. The procedure will be modified to adapt to a testee's recall on previous screen presentations in order to obtain precise quantitative estimates of individual differences in visual sensory memory capacity with fewer replications. For example, the size of the array (and thus the memory capacity demands) can be made larger or smaller based on an individual's previous performance; and/or the duration of the array presentation could be lengthened or shortened to adapt the task difficulty to the testee's ability level during the course of testing. Prior to implementing such an adaptive approach, however, psychometric research is needed to determine how much of an increase or decrease in stimulus array size constitutes a meaningful increment or decrement in difficulty and what ranges of array size are needed to adequately span differences existing in various populations of interest.

Data collection on two tests of short-term spatial and perceptual memory was therefore designed to allow a preliminary evaluation of potential adaptive testing parameters. The experimental Memory for Patterns items consisted of bounded two-dimensional arrays containing 3 to 10 letters. Each item consisted of a pair of successive screen presentations. The stimulus display was presented for a brief timed period and then was erased from the cathode ray terminal (CRT) and replaced with the recall display. The recall display contained a bounded letter pattern that was identical to the first pattern (the stimulus display) except that one or two letters had changed position. The recall display was untimed and accepted the student's response, indicating which letter(s) she/he thought had moved. Figure 3 shows sample Memory for Patterns stimulus and response display, constituting one test item. Data were also collected on a related set of items designed to measure Space Memory, which were similar to the Memory for Patterns items except that the letter patterns presented were unbounded and thus spread about the entire CRT screen.

In the initial data collection for the Memory for Patterns and Space Memory items, subtests varied in the number of letters that could change position from the first pattern to the second. Three 10-item subtests were administered to each student in a number of different experimental groups. The first subtest was composed of patterns in which one letter changed position, the second subtest was composed of patterns in which two letters changed position, and the third subtest included patterns in which either one or two letters changed position. Students were assigned to one of four conditions that varied the order in which the items were presented and the presentation time, in seconds, of the first pattern of each item pair. In the Memory for Patterns test the two presentation times were 5 and 10 seconds, and in the Space Memory test they were 7 and 12 seconds. The two orders in which items were presented were (1) from lowest to highest difficulty and (2) from moderate to highest to lowest difficulty,

Figure 3
A Sample Memory for Patterns Item, with Stimulus Display
on the Left and Recall Display on the Right



where difficulty was indexed by the number of letters in the pattern. Reactions to the tests were also obtained from each student.

The analysis of the data was directed at scaling the difficulty of items and at studying the effects of pattern density and presentation times on item difficulty indices. A comparison of the item order conditions was made to examine the effects of practice and proactive inhibition on indices of item difficulty. A comparison of presentation time conditions was directed at determining reasonable item exposure times and at investigating the possibility of using presentation time along with pattern density and number of pattern changes as adaptive parameters for future adaptive testing.

The results of preliminary analyses of the Memory for Patterns items were used to design new items and to modify the experimental conditions under which they were administered. The major conclusions from the preliminary analysis was that item difficulty was more a function of type of pattern configuration than either rate or order condition. For this reason, types of Memory for Patterns items were hypothesized in a systematic manner. Eight Memory for Patterns item types, which can be separated into two groups, were developed. One group of items is composed of patterns taking a geometric form, such as a line, triangle, square, or pentagon. Four item types of this nature were developed:

1. An item that had a geometric form was changed to a nongeometric form through a small move,
2. An item that had a geometric form was changed to a nongeometric form through a large move,
3. An item that had a nongeometric form was changed into a geometric form through a small move, and
4. An item that had a nongeometric form was changed into a geometric form through a large move.

Within the four item types, there were various degrees of nongeometric form that the patterns could have.

A second group of items was composed of nongeometric forms, but the patterns varied in terms of pattern configuration definition. For example, some patterns, although not geometric in form, are better defined and thus easier to

remember. The four item types were as follows:

1. Well-defined pattern configuration with a small change in letter configuration,
2. Well-defined pattern configuration with a large change in letter configuration,
3. Poorly defined pattern configuration with a small change in letter position, and
4. Poorly defined pattern configuration with a large change in letter position.

Hypotheses were made with regard to item type and resultant item difficulty. Ninety new items were written based on these eight basic Memory for Patterns item types.

Of the 90 items, 42 were administered under various experimental conditions. Students were assigned sequentially to one of 15 testing conditions. The 42 Memory for Patterns items were presented under three order and five rate conditions. Values of pattern densities, or the number of letters in a configuration were 3, 4, 5, 6, 7, 8, 9, and 10 letters. The 42 items were presented in three orders: (1) ordered from 3 to 10 letters in a pattern; (2) ordered from 6 to 10 letters, then 3 to 5 letters in a pattern; and (3) ordered from 8 to 10, then 3 to 7 letters in a pattern. The five rate conditions were 3, 5, 7, 10, and 13 seconds in duration. Analyses of these data will be oriented toward the identification of adaptive parameters for the items and the effects of administration conditions (e.g., sequence effects) on the adaptive parameters.

Digit Span. Contrasting with the Memory for Patterns tests, which tap both spatial abilities and short-term memory capacity, the Digit-Span test is primarily a test of short-term memory. In this test a series of numbers is presented in rapid succession. The respondent is asked to recall the numbers in the order they were presented by typing them into the computer terminal in a serial string. Since the time interval between presenting the numbers, clearing the screen, and asking for a response was very short, the test was essentially an indicator of short-term memory capacity.

To identify possible adaptive parameters for this type of test, presentation rate was varied experimentally so that items were presented at one of three rates--.2 seconds, .3 seconds, and .5 seconds--corresponding to fast, moderate, and slow rates. Series length (the number of stimulus values to be recalled), a second potential adaptive testing parameter, varied from 4 to 10 stimuli. Six items at each test length were administered to yield a total of 42 digit series in the test. Length of digit series and of presentation rate will be investigated as possible test and item parameters.

The items were presented in three different orders: (1) from easy to difficult (where difficulty was defined in terms of numbers of stimulus values to recall), (2) from moderately difficult to difficult to easy, and (3) from difficult to easy to moderately difficult. The order of item presentation will be analyzed to determine if there are practice or prohibitive effects from one item to the next, since such effects would be undesirable in adaptive test administration, and to determine the effects of series length and display time on item difficulties.

Analyses and Future Plans

Data analyses and future data collection will be oriented toward identifying potential adaptive parameters in terms of factors influencing the difficulty of test items. In addition, the influence of undesirable factors--such as sequence effects, inhibitive effects, and other factors that would interfere with adaptive administration--will be investigated in order to permit the design and evaluation of adaptive tests of information-processing abilities.

Other measures of more traditional psychometric factors, such as flexibility of closure, may also benefit from application of a computerized adaptive testing framework. A common measure of this construct has been variants of the Hidden or Embedded Figures test. The unique capabilities of computerized administration may allow increased validity to be achieved by inducing movement into either the stem figure and/or the alternative response figures. For example, the testee may be required to selectively attend to and to articulate the stem figure in a more complex figure, which not only contains distracting lines but also grows, shrinks, translates, and/or rotates over time. Adaptive administration can be achieved by modifying the amount of "noise" in the figure and by dynamically adapting the amount and speed of movement in the complex figure. Several psychometric questions can be studied. For example, what size increments in "noise" and movement will allow the computer to most efficiently converge upon the testee's ability level? Are amounts of "noise" and movement independent dimensions of difficulty to be manipulated? How is performance under varying degrees of noise and movement to be scored?

The design and implementation of adaptive tests of these information-processing kinds of ability tests raises a host of new questions and problems to be investigated. Beyond the identification of adaptive parameters for these kinds of items, and ruling out extraneous factors such as sequence effects (which can reduce the effectiveness of the adaptive procedures), new questions will need to be addressed with regard to the design and scoring of the adaptive tests. Design questions will include the identification of the functions relating display time to item difficulty and identification of the procedures for using this continuous function in adapting display time on each item to each individual's test performance on an item-by-item basis. Since it may be observed that this function is different for items of different difficulties based on stimulus characteristics (e.g., pattern density, length of span string), adaptive testing procedures will need to be developed that will jointly take into account the combination of discrete and continuous difficulty factors. New scoring procedures may also need to be designed for these kinds of tests, since each testee will receive a set of items selected to match his/her ability levels; and/or the applicability of IRT-based scoring procedures will need to be investigated.

Finally, comparisons of computerized adaptive, computerized nonadaptive and traditional measures of these abilities should be made to determine if more precise and efficient measurement can be achieved through adaptive administration. Where appropriate external criteria are available, predictive validity comparisons should also be made. If the findings from traditional ability testing generalize to these kinds of ability tests, the resulting tests will be shorter, more precise, and more valid and will permit more meaningful measurement of the range of human abilities.

References

Bejar, I. I. An investigation of the dichotomous, graded, and continuous response level latent trait models. Unpublished doctoral dissertation, University of Minnesota, 1975.

Bejar, I. I. Assessing the unidimensionality of achievement tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Bejar, I. I. A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 1980, 17, 283-296.

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977.

Carroll, J. Pyschometric tests as cognitive tasks: A new "structure of intellect." In L. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Laurence Erlbaum/Halsted Press, 1976.

Chiang, A., & Atkinson, R. C. Individual differences and interrelationships among a select set of cognitive skills. Memory and Cognition, 1976, 4, 661-672.

Cory, C. H. Relative utility of computerized versus paper-and pencil tests for predicting job performance. Applied Psychological Measurement, 1977, 1, 551-564.

Cory, C. H., Rimland, B., & Bryson, R. A. Using computerized tests to measure new dimensions of abilities: An exploratory study. Applied Psychological Measurement, 1977, 1, 101-110.

Day, R. S. Systematic individual differences in informtion processing. In P. G. Zimbardo and F. L. Ruch (Ed.), Psychology and Life. Glenview, IL: Scott & Foresman, 1977.

Donlon, T. F., & Fischer, F. E. An index of an individual's agreement with group-determined item difficulties. Educational and Psychological Measurement, 1968, 28, 105-113.

French, J. W., Ekstrom, R. B., & Price, L. A. Kit of reference tests for cognitive factors. Princeton, NJ: Educational Testing Service, 1963.

Hunt, E. B., Frost, N., & Lunneborg, C. E. Individual differences in cognition. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 7). New York: Academic Press, 1973.

Hunt, E. B., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

Jacobs, P. I. A study of large score changes on the SAT (Research Bulletin RB 63-20). Princeton, NJ: Educational Testing Service, June 1963.

Johnson, M. F., & Weiss, D. J. Parallel forms reliability and measurement accuracy comparison of adaptive and conventional testing strategies. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, 1980.

Kingsbury, G. G., & Weiss, D. J. An alternate-forms reliability and concurrent validity comparison of Bayesian adaptive and conventional ability tests (Research Report 80-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1980.

Levine, M., & Drasgow, F. Appropriateness measurement: Basic principles and validating studies. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, 1980.

Levine, M. V., & Rubin, D. B. Measuring the appropriateness of multiple-choice test scores. Journal of Educational Statistics, 1979, 4, 269-290.

Lunneborg, C. E. Information-processing correlates of the Raven Progressive Matrices Tests (Technical Report 77-10). Seattle, WA: University of Washington, Educational Assessment Center, 1977.

McBride, J. R. Adaptive verbal ability testing in a military setting. In D. J. Weiss (Ed.), Proceedings of the 1979 Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, 1980.

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976.

Reckase, M. D. The effect of item pool characteristics on the operation of a tailored testing procedure. Paper presented at the spring meeting of the Psychometric Society, Murray Hill, NJ, April 1976.

Reckase, M. D., Ability estimation and item calibration using the one and three parameter logistic models: A comparative study (Research Report 77-1). Columbia: University of Missouri, Educational Psychology Department, Tailored Testing Research Laboratory, November 1977.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Rose, A. Human information processing: An assesment and research battery (Technical Report 46). Ann Arbor, MI: University of Michigan, Human Performance Center, 1974.

Sperling, G. The information available in brief visual presentations. Psychological Monographs, 1960, 74, (Whole No. 498).

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Trabin, T. E., & Weiss, D. J. The person response curve: Fit of individuals to item characteristic curve models (Research Report 79-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1979.

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

Urry, V. W. Individualized testing by Bayesian estimation (Research Bulletin 0171-177). Seattle: University of Washington, Bureau of Testing, April 1971.

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1975. (NTIS No. AD A020961)

Vale, C. D., & Weiss, D. J. A comparison of information functions of multiple-choice and free-response vocabulary items (Research Report 77-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1977.

Wood, R. L., & Lord, F. M. A user's guide to LOGIST. (Research Memorandum RM-76-4). Princeton, NJ: Educational Testing Service, May 1976.

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

Wright, B. Solving measurement problems with the Rasch Model. Journal of Educational Measurement, 1977, 14, 97-115.

Distribution List

Navy

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

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

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

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

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

1 Dr. Henry M. Halff
Department of Psychology,C-009
University of California at San Diego
La Jolla, CA 92093

1 LT Steven D. Harris, MSC, USN
Code 6021
Naval Air Development Center
Warminster, Pennsylvania 18974

1 CDR Charles W. Hutchins
Naval Air Systems Command Hq
AIR-340F
Navy Department
Washington, DC 20361

1 CDR Robert S. Kennedy
Head, Human Performance Sciences
Naval Aerospace Medical Research Lab
Box 29407
New Orleans, LA 70189

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

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

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock
Newport News, VA 23607

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

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

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

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

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

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

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

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

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

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

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 Dr. Donald F. Parker
Graduate School of Business Administrati
University of Michigan
Ann Arbor, MI 48109

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland (03B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

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

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

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

1 Roger Weissinger-Baylon
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr John H. Wolfe
Code P310
U. S. Navy Personnel Research and
Development Center
San Diego, CA 92152

Army

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

1 Dr. Dexter Fletcher
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria,VA 22333

1 DR. FRANK J. HARRIS
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Col Frank Hart
Army Research Institute for the
Behavioral & Social Sciences
5001 Eisenhower Blvd.
Alexandria, VA 22333

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

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

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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

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

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

Air Force

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

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

1 Dr. Marty Rockway
Technical Director
AFHRL(OT)
Williams AFB, AZ 58224

2 3700 TCHTW/TTGH Stop 32
Sheppard AFB, TX 76311

Marines

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

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

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

Other DoD

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

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

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

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

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

1 Dr. Joseph I. Lipson
SEDR W-638
National Science Foundation
Washington, DC 20550

1 William J. McLaurin
Rm. 301, Internal Revenue Service
2221 Jefferson Davis Highway
Arlington, VA 22202

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

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

1 Dr. Frank Withrow
U. S. Office of Education
400 Maryland Ave. SW
Washington, DC 20202

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

Non Govt

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Anderson, Thomas H., Ph.D.
Center for the Study of Reading
174 Children's Research Center
51 Gerty Drive
Champiagn, IL 61820

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

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

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

1 Dr. Patricia Baggett
Department of Psychology
University of Denver
University Park
Denver, CO 80208

1 Mr Avron Barr
Department of Computer Science
Stanford University
Stanford, CA 94305

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

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80309

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 Dr. Bruce Buchanan
Department of Computer Science
Stanford University
Stanford, CA 94305

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

1 Dr. Pat Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

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

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

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

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

1 Dr. Lynn A. Cooper
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Kenneth B. Cross
Anacapa Sciences, Inc.
P.O. Drawer Q
Santa Barbara, CA 93102

1 Dr. Ronna Dillon
Department of Guidance and Educational P
Southern Illinois University
Carbondale, IL 62901

1 Dr. Hubert Dreyfus
Department of Philosophy
University of California
Berkely, CA 94720

1 LCOL J. C. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARC
NATIONAL DEFENCE HQ
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A OK2

1 Dr. Ed Feigenbaum
Department of Computer Science
Stanford University
Stanford, CA 94305

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

1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

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

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

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9

1 Dr. R. Edward Geiselman
Department of Psychology
University of California
Los Angeles, CA 90024

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

1 Dr. Marvin D. Glock
217 Stone Hall
Cornell University
Ithaca, NY 14853

1 Dr. Daniel Gopher
Industrial & Management Engineering
Technion-Israel Institute of Technology
Haifa
ISRAEL

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Glenda Greenwald, Ed.
"Human Intelligence Newsletter"
P. O. Box 1163
Birmingham, MI 48012

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

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Kenneth A. Klivington
Program Officer
Alfred P. Sloan Foundation
630 Fifth Avenue
New York, NY 10111

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

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

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

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

1 Dr. Robert A. Levit
Director, Behavioral Sciences
The BDM Corporation
7915 Jones Branch Drive
McClean, VA 22101

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

1 Dr. Erik McWilliams
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Dr. Mark Miller
Computer Science Laboratory
Texas Instruments, Inc.
Mail Station 371, P.O. Box 225936
Dallas, TX 75265

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

1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093

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

1 Dr. Seymour A. Papert
Massachusetts Institute of Technology
Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139

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

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

1 Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80302

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80309

1 Dr. Steven E. Poltrock
Department of Psychology
University of Denver
Denver,CO 80208

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

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

1 Dr. Lauren Resnick
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

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

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

1 Dr. David Rumelhart
Center for Human Information Processi
Univ. of California, San Diego
La Jolla, CA 92093

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 Dr. Alan Schoenfeld
Department of Mathematics
Hamilton College
Clinton, NY 13323

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

1 Committee on Cognitive Research
% Dr. Lonnie R. Sherrod
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Robert S. Siegler
Associate Professor
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

1 Dr. Edward E. Smith
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

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

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

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

1 David E. Stone, Ph.D.
Hazeltine Corporation
7680 Old Springhouse Road
McLean, VA 22102

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

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

1 Dr. John Thomas
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
1845 S. Elena Ave.
Redondo Beach, CA 90277

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

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

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

1 DR. GERSHON WELTMAN
PERCEPTRONICS INC.
6271 VARIEL AVE.
WOODLAND HILLS, CA 91367

1 Dr. Keith T. Wescourt
Information Sciences Dept.
The Rand Corporation
1700 Main St.

Previous Publications (continued)

77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-4. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

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

Computerized Adaptive Testing Laboratory
N660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis MN 55455 U.S.A.

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

81-2. Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing. February 1981.

81-1. Review of Test Theory and Methods. January 1981.

80-5. An Alternate-Forms Reliability and Concurrent Validity Comparison of Bayesian Adaptive and Conventional Ability Tests. December 1980.

80-4. A Comparison of Adaptive, Sequential, and Conventional Testing Strategies for Mastery Decisions. November 1980.

80-3. Criterion-Related Validity of Adaptive Testing Strategies. June 1980.

80-2. Interactive Computer Administration of a Spatial Reasoning Test. April 1980.

Final Report: Computerized Adaptive Performance Evaluation. February 1980.

80-1. Effects of Immediate Knowledge of Results on Achievement Test Performance and Test Dimensionality. January 1980.

79-7. The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models. December 1979.

79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.

79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.

79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.

79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.

Final Report: Bias-Free Computerized Testing. March 1979.

79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.

79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.

78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.

78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.

78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.

78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.

78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.

77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.

77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.

77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.

77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.

-continued overleaf-