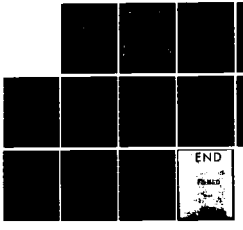
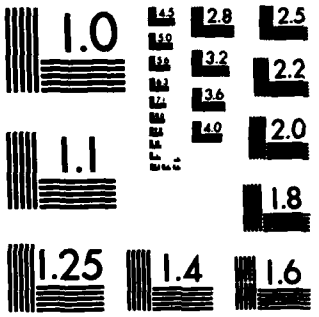
AD-A144 065 COMPUTER-BASED MEASUREMENT OF INTELLECTUAL CAPABILITIES 1/1
(U) MINNESOTA UNIV MINNEAPOLIS DEPT OF PSYCHOLOGY
D J WEISS DEC 83 N00014-76-C-0243

UNCLASSIFIED F/G 9/2 NL





(12)

Final Report

Computer-Based Measurement of Intellectual Capabilities

David J. Weiss

December 1983

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

FINAL REPORT OF PROJECT NR150-382, N00014-76-C-0243
SUPPORTED BY THE
PERSONNEL AND TRAINING RESEARCH PROGRAMS
PSYCHOLOGICAL SCIENCES DIVISION
OFFICE OF NAVAL RESEARCH
AND THE
NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. ADA144065	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Final Report: Computer-Based Measurement of Intellectual Capabilities		5. TYPE OF REPORT & PERIOD COVERED Final Report. September 1976-January 1983
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0243
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61553N Proj.:RR042-04 T.A.: RR042-04-01 W.U.: NR150-382
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington VA 22217		12. REPORT DATE December 1983
		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

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

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

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Navy Personnel Research and Development Center and the Office of Naval Research, and monitored by the Office of Naval Research.

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

Adaptive Testing	Response-Contingent Testing
Computerized Testing	Latent Trait Theory
Tailored Testing	Response Latency
Individualized Testing	Psychological Reactions to Testing

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

The research program's objectives are described, and the research approach is summarized and related to the sixteen technical reports completed under this contract. Fifteen major research findings are presented. The implications of the research findings and methods for future research in computerized testing and adaptive testing are described. Also included are abstracts of the sixteen technical reports.

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)

CONTENTS

FINAL REPORT:
COMPUTER-BASED MEASUREMENT OF INTELLECTUAL CAPABILITIES

Objectives

The objectives of this research program were based on a review of previous research literature that identified the potential of computerized adaptive testing to reduce at least five kinds of errors in the measurement of human capacities:

1. Errors due to mismatch of test item difficulty with testee ability;
2. Errors due to the psychological effects of testing;
3. Errors due to inappropriate dimensionality;
4. Errors due to failure to extract sufficient information from the testee;
5. Errors due to over-simplistic conceptualizations of intellectual capabilities.

Within the context of these five sources of error, which act to reduce the precision, accuracy and utility of current ability testing procedures, the research was designed to:

1. Extend previous research efforts to identify the most useful computer-based adaptive testing strategies.
2. Study the psychological effects of computerized adaptive testing, to identify those testing conditions which minimize adverse effects and maximize positive effects.
3. Investigate the problem of intra-individual multidimensionality in ability testing.
4. Examine the use of such response modes as probabilistic responding and free-response methods for use in computerized adaptive testing in order to extract maximum information from each examinee's response to each test item.
5. Develop, refine and evaluate new computer-administered ability tests which measure abilities not now measurable using paper and pencil ability testing.

Research in pursuance of these primary objectives began in September 1975 and continued through December 1978. A contract extension, funded by the Navy Personnel Research and Development Center, was designed to complete a live-testing validity comparison of adaptive and conventional tests using Marine recruits. This extension continued the contract through September 1979. Technical reports were completed through January 1983.

Approach

The major focus of the research was on the evaluation of adaptive testing strategies by comparison of their characteristics with each other and with conventional tests. Both monte carlo simulation and live testing were used in these studies. In Research Report 75-6 the stradaptive testing strategy was examined in monte carlo simulation to evaluate various scoring techniques possible with this testing strategy, under various test lengths and prior information conditions. Performance of the stradaptive testing strategy was also evaluated in live testing (Research Report 80-3) by comparing its validity with that of a conventional test and a Bayesian adaptive test.

The Bayesian adaptive testing strategy was further studied in several reports. Monte carlo simulation was used in Research Report 76-1 to examine the performance of this testing strategy under several item pool configurations and at a number of test lengths. In Research Reports 80-5 and 83-1, the reliability and validity of the Bayesian adaptive test was compared with that of conventional tests in a college population (80-5) and in a military recruit population (83-1). Research Report 77-4 describes a procedure for improving the efficiency of item selection in Bayesian adaptive testing.

Several other problems concerned with the application of adaptive tests to the measurement of abilities were discussed in a symposium presented at the 1976 meeting of the Military Testing Association (Research Report 77-1). An overview of adaptive testing strategies, presented by McBride, included a discussion of item selection strategies, scoring adaptive tests, and problems of evaluating adaptive tests. The problem of estimating trait status in adaptive testing based on item response theory approaches was presented by Sympson, including a comparison of the characteristics of Bayesian and likelihood-based estimates. Vale, in his paper, considered the problem of classifying individuals into discrete ability categories (e.g., pass-fail); his monte carlo analysis compared adaptive and conventional tests designed for making dichotomous classifications.

The effects of testing conditions on test performance were investigated in a number of live-testing studies. Since computer-administered testing permits immediate scoring of an examinee's answer to a test question, it becomes possible to inform the examinee immediately after each response is given as to whether the answer was correct or incorrect. This immediate knowledge of results, or immediate feedback, was investigated in several studies in terms of its effects on ability test performance in adaptive and conventional tests (Research Reports 76-3 and 78-2), its interaction with test difficulty (Research Report 78-2) and computer versus self-paced test administration (Research Report 81-2), and its effects on examinees' reactions to test administration (Research Reports 76-4 and 81-2). Related studies examined the effects of time limits on test-taking behavior (Research Report 76-2) and the accuracy of the perceived difficulty of test items (Research Report 77-3).

The question of intra-individual dimensionality in performance on ability tests was recast within the more general framework of the fit of individuals to item response theory (IRT) models. This issue was examined in one study (Research Report 79-7) in which the predicted and acutal performance of single individuals was examined for indications of lack of person fit due to intra-indi-

vidual multidimensionality or other factors reflecting non-fit to the unidimensional IRT models.

The use of test item response modes other than the multiple-choice item was examined in one study (Research Report 77-2) which compared test information derived from free-response administration to that of the same items administered in multiple-choice mode.

The use of the unique capability of interactive computers to measure abilities not measurable by paper-and-pencil tests was examined in one study (Research Report 80-2). An interactive spatial reasoning test was designed based on the popular "15 puzzle" in which examinees were required to restructure a set of 15 numerals into a target pattern using a minimum number of moves. Examinee performance on the test was analyzed in terms of such factors as number of moves to solution, quality of the moves, and response latencies at each point in the testing procedure.

Major Findings

The major findings below are generally organized according to the original objectives of the research program. Additional details are in the Research Report abstracts. Many of the original Research Reports contain additional important findigs.

Adaptive Testing Strategies

1. Monte carlo data comparing the stradaptive test with non-adaptive approaches to ability testing (Research Report 75-6) shows that the stradaptive test provides more equiprecise measurement than a peaked conventional test. As item discriminations increased, the equiprecision of the stradaptive test increased relative to that of the conventional test.

2. A stradaptive test with an average of 25% fewer items than a conventional test obtained significantly higher validities with a college grade-point average criterion than did the conventional test (Research Report 80-3).

3. Monte carlo evaluation of a Bayesian adaptive testing strategy identified a number of psychometric problems in the ability estimates resulting from this testing strategy (Research Report 76-1). Bayesian ability estimates were highly correlated with test length, were non-linearly biased for about two-thirds of the ability range, and were dependent on the prior ability estimate.

4. Although the monte carlo simulations of the Bayesian adaptive test identified these potential problems with the Bayesian ability estimates, they appeared to have little impact on the reliability and validity of Bayesian ability estimates. Live-testing studies of the Bayesian adaptive testing strategy in a college population showed validities equal to that a conventional test (Research Report 80-3), and high reliabilities for tests of 2 to 30 items in length (Research Report 80-5); in the latter study, however, using a concurrent validity criterion, the conventional test had higher validity correlations than the adaptive test. In a military recruit popu-

lation (Research Report 83-1), the Bayesian adaptive test achieved both higher validities and higher reliabilities than did a comparable conventional test. In this population, a 9-item adaptive test achieved the same reliability as a 17-item conventional test; 10- to 11-item adaptive tests achieved the same concurrent validities as 28- to 30-item conventional tests.

5. The original form of the Bayesian adaptive test used an item-search procedure that could require excessive amounts of computing time for an interactive test administration environment. A rapid item-search procedure was developed and shown to select the same subset of items as the original procedure in about one-tenth the amount of computer time.

6. Different methods of estimating ability from adaptive tests have different characteristics. Validities in the prediction of college grade-point averages from a stradaptive test were higher for ability estimates not based on IRT methods than they were for IRT-based ability estimates (Research Report 80-3). Within the IRT methods for estimating ability, Bayesian methods are slightly order dependent, resulting in slightly different ability estimates with the same items administered in different orders (Sympson, in Research Report 77-1). Bayesian ability estimates also have different psychometric characteristics than do estimates based on maximum likelihood procedures.

7. Adaptive tests can be used for classification purposes as well as for measurement on a continuous scale. When compared to conventional tests designed to make classifications, adaptive tests can classify more accurately than conventional tests when it is necessary to make more than a single dichotomous classification based on test scores (Vale, in Research Report 77-1).

Test Administration Conditions

8. An analysis of response latency data showed that testees approach different testing procedures in different ways (Research Report 76-2). The response latency data suggest that these different test-taking styles and strategies might be potentially useful as moderator or predictor variables in the prediction of external criteria.

9. Computer-administered feedback (immediate knowledge of results) on a conventional test appears to result in enhanced ability test performance for testeees of all ability levels (Research Report 76-3). Under computer-administered feedback conditions, mean test scores were significantly higher for both high- and low-ability testees. Ninety percent of college students favorably evaluated their experience with computer-administered feedback (Research Report 76-4).

10. Adaptive tests appear to be more intrinsically motivating for low-ability testees (Research Report 76-4), and result in higher ability estimates (Research Report 76-3), than similarly administered conventional tests. This suggests that adaptive testing might eliminate some of the undesirable psychological effects characteristic of conventional testing procedures, resulting in fairer and more accurate test scores for testees who typically

obtain low scores on conventional ability tests.

11. Item-difficulty perceptions of college students were highly related to objective indices of test item difficulty (Research Report 77-3). This suggests that test difficulty, which may differ between conventional and adaptive tests for examinees of the same ability, might be an important factor affecting the test performance of individuals.

12. Test difficulty interacted with immediate knowledge of results to produce effects on ability estimates, but not on psychological reactions to the testing conditions (Research Report 78-2). Since difficulty is more equal across ability levels in an adaptive test than in a conventional test, these results suggest that the testing environment of adaptive tests will result in fewer sources of error in ability estimates than will conventional ability tests.

Other Findings

13. Analysis of person-fit data derived from the person response curve indicated that the vast majority of college students studied responded to a set of test items in accordance with the 3-parameter logistic IRT model (Research Report 79-7). The person response curve approach also identified a small group of individuals whose responses to the test items appeared to result from an underlying multidimensional ability structure with respect to the ability domain studied.

14. The dependence of adaptive testing on the multiple-choice item will result in test scores with less than optimal properties. Analysis of free-response item data indicates that more informative ability estimates can be derived from free response items than from the same items administered as multiple-choice items and scored by optimal IRT methods; differences were greater for high-ability examinees (Research Report 77-2).

15. Interactive computer administration of ability test items permits the design and implementation of ability tests using novel item formats, which may extend the range of measurable abilities beyong those now measurable using a dimensional approach. The design and implementation of an interactive spatial problem-solving test (Research Report 80-2) permitted the measurement and analysis of a number of problem-solving types of variables that described individual differences in problem-solving styles; these variables might be useful as ability kinds of variables, following further study and refinement.

Implications for Further Research

The findings and experience of this research program support the feasibility, utility and psychometric advantages of computerized adaptive measurement of intellectual capabilities. However, many new questions were raised by the research, and some of the original questions addressed are still in need of further research.

Adaptive Testing Strategies

Research has concentrated on comparison of the stradaptive and Bayesian adaptive testing strategies with conventional tests. Further research is needed (1) comparing these strategies directly with each other, in both live testing and in simulation, and (2) in comparing these strategies with other adaptive testing strategies, such as an information-based item selection routine.

All adaptive testing strategy comparisons to date that used monte carlo simulation techniques have made two assumptions that are not characteristic of real data. First, they have assumed that the item pool is characterized by items with known parameter values. In real item pools, however, item parameter values are never known, but are always estimated. These estimates are only approximations to the true values and, as a consequence, contain some degree of error, with rather substantial degrees of error for some of the item parameters. Since adaptive testing strategies are designed to explicitly select items based on these item parameter estimates, the possibility exists that in a real item pool with error-laden item parameters adaptive tests might perform less optimally due to the error in the item parameter estimates. Thus, simulation studies should be designed and implemented to experimentally vary the degrees of error in item parameter estimates and to evaluate the effects of these errors on the performance of adaptive testing strategies, in order to identify the effects of these errors on the performance of the testing strategies.

A second assumption made in all monte carlo comparisons of adaptive testing strategies is that the item pool is strictly unidimensional, since only one set of item parameter values is used for each item. In real data, however, item pools are very rarely strictly unidimensional. Frequently, item pools are characterized by second and succeeding factors that account for from trivial portions of the item pool variance to substantial portions of that variance. While multidimensional IRT models have not yet been sufficiently operationalized to permit the estimation of item parameters for dimensions beyond the first, it is possible to examine the effects of multidimensionality on adaptive testing strategies. One approach to studying this problem is to simulate the administration of adaptive testing strategies with unidimensional item parameters when item responses are generated from an underlying multidimensional structure. This approach assumes that the dimensionality of the item responses is the true underlying multidimensional structure, while the apparent unidimensionality of the item pool is the result of the item parameterization process applied to it. Studies of this type would enable the identification of the degrees and types of multidimensionality that could be tolerated by the various adaptive testing strategies without serious degradation of their performance.

Further live-testing comparisons of adaptive testing strategies are also necessary. The four live-testing studies completed under this contract yielded somewhat conflicting results. In two of the four studies, adaptive tests obtained higher validities than conventional tests with a smaller average number of items, and in one study with a smaller median number of items. In the study using military recruits a very clear advantage was obvious for the adaptive tests beginning at short test lengths. When a large group of college students was studied, however, although the expected differences in reliability were obtained, the conventional test performed better on the concurrent validity crite-

rion. Since the design of the two large-sample studies was similar, differences in results could be attributable to differences in the examinees, the item pools, or the criterion tests. Additional live-testing studies are needed to evaluate the effects of these conditions, as well as to evaluate the performance of other adaptive testing strategies and to evaluate their performance with additional criterion variables.

Test Administration Conditions

The research results show that a number of test administration variables influence test scores, IRT-based ability estimates, and/or examinees' reactions to tests. These include test speededness, test difficulty, and immediate feedback to examinees as to whether their item responses are correct or incorrect. Testing strategy (adaptive versus conventional) also had some effects on test performance and reactions, probably due to the differing difficulties of adaptive and conventional tests. Immediate feedback of results appeared to be an important potential factor in increasing test-taking motivation and improving test scores.

Studies completed on the effects of test administration conditions have all utilized volunteer college students as examinees and have used verbal ability items in the tests administered. Since the test-taking motivation of volunteer students might differ when tested under conditions where the tests are being used for grading or other purposes, future studies should examine the effects of test administration conditions when the tests being administered are to be used for purposes other than research. In addition, the generality of the observed effects should be studied on populations other than college students, and using other tests in addition to verbal ability tests. Further studies should also include the effects of other adaptive testing strategies as test administration conditions, in conjunction with immediate knowledge of results.

Intra-Individual Dimensionality, Response Modes, and New Abilities

Research in these three areas was only begun during the contract period. The person characteristic curve results show that the vast majority of the one group of college students studied responded to a set of test items in accordance with the three-parameter logistic IRT model. A small group of students was identified, however, whose responses appeared to be reliably divergent from that model. These deviations were ascribed to intra-individual multidimensionality. Since the person response curve method was used in only this one study, further studies are indicated. Of importance is the performance in monte carlo simulations of the person-fit indices under conditions of unidimensionality, the derivation of appropriate sampling distributions of the person-fit indices, the evaluation of alternate person-fit indices, and the effect of test structure characteristics (e.g., distributions of item characteristics) on the performance of person-fit indices. Additional live-testing studies should also be implemented to study the effects of various test administration conditions (e.g., interruptions, poor testing conditions, immediate knowledge of results) on intraindividual dimensionality by means of the person response curve and associated indices of person fit.

Failure to extract sufficient information from an examinee's responses to

multiple-choice test items can lower the quality of obtained measurements. The one study completed on this problem indicated that the use of free-response items was able to improve the measurement precision of a set of vocabulary items beyond that possible from scoring the same items as polychotomous multiple-choice items. Both of these administration/scoring modes provide better measurement than dichotomously-scored multiple-choice items. Since this study used college students on a single short vocabulary test, further studies are obviously needed to examine the generality of the results. In addition, research is needed to examine the performance of other alternatives to the dichotomously-scored multiple-choice item such as probabilistic responding, which are now feasible when administered by interactive computers.

Interactive computer administration of ability tests makes possible the development of a wide range of new kinds of ability tests to supplement the standard dimensionality-based tests currently in use. This project has demonstrated that interactive administration of a problem-solving type of test can result in substantial amounts of new kinds of data on examinees in addition to the traditional number of items answered correctly. These data can include information on problem-solving styles and response latencies that might be indicative of other individual differences problem-solving variables. Future research should investigate the psychometric characteristics of these variables including their reliabilities and their contributions to validity, as well as examine the utility of the interactive computer for measuring other abilities such as spatial, perceptual, and memory abilities which are now possible to be measured by computer administration.

RESEARCH REPORT ABSTRACTS

Research Report 75-6
A Simulation Study of Stradaptive Ability Testing
C. David Vale and David J. Weiss
December 1975

A conventional test and two forms of a stradaptive test were administered to thousands of simulated subjects by minicomputer. Characteristics of the three tests using several scoring techniques were investigated while varying the discriminating power of the items, the lengths of the tests, and the availability of prior information about the testee's ability level. The tests were evaluated in terms of their correlations with underlying ability, the amount of information they provided about ability, and the equiprecision of measurement they exhibited. Major findings were (1) scores on the conventional test correlated progressively less with ability as item discriminating power was increased beyond $\theta = 1.0$; (2) the conventional test provided increasingly poorer equiprecision of measurement as items became more discriminating; (3) these undesirable characteristics were not characteristic of scores on the stradaptive test; (4) the stradaptive test provided higher score-ability correlations than the conventional test when item discriminations were high; (5) the stradaptive test provided more information and better equiprecision of measurement than the conventional test when test lengths and item discriminations were the same for the two strategies; (6) the use of valid prior ability estimates by stradaptive strategies resulted in scores which had better measurement characteristics than scores derived from a fixed entry point; (7) a Bayesian scoring technique implemented within the stradaptive testing strategy provided scores with good measurement characteristics; and (8) further research is necessary to develop improved flexible termination criteria for the stradaptive test. (AD A020961)

Research Report 76-1
Some Properties of a Bayesian Adaptive Ability Testing Strategy
James R. McBride and David J. Weiss
March 1976

Four monte carlo simulation studies of Owen's Bayesian sequential procedure for adaptive mental testing were conducted. Whereas previous simulation studies of this procedure have concentrated on evaluating it in terms of the correlation of its test scores with simulated ability in a normal population, these four studies explored a number of additional properties, both in a normally distributed population and in a distribution-free context. Study 1 replicated previous studies with finite item pools, but examined such properties as the bias of estimate, mean absolute error, and correlation of test length with ability. Studies 2 and 3 examined the same variables in a number of hypothetical infinite item pools, investigating the effects of item discriminating power, guessing, and variable vs. fixed test length. Study 4 investigated some properties of the Bayesian test scores as latent trait estimators, under three different item pool configurations (regressions of item discrimination on item difficulty). The properties of interest included the regression of latent trait estimates on actual trait levels, the conditional bias of such estimates, the information curve

of the trait estimates, and the relationship of test length to ability level. The results of these studies indicated that the ability estimates derived from the Bayesian test strategy were highly correlated with ability level. However, the ability estimates were also highly correlated with number of items administered, were non-linearly biased, and provided measurements which were not of equal precision at all levels of ability. (AD A022964)

Research Report 76-2
Effects of Time Limits on Test-Taking Behavior
T. W. Miller and David J. Weiss
April 1976

Three related experimental studies analyzed rate and accuracy of test response under time-limit and no-time-limit conditions. Test instructions and multiple-choice vocabulary items were administered by computer. Student volunteers received monetary rewards under both testing conditions. In the first study, college students were blocked into high- and low-ability groups on the basis of pretest scores. Results for both ability groups showed higher response rates under time-limit conditions than under no-time-limit conditions. There were no significant differences between the time-limit and no-time-limit accuracy scores. Similar results were obtained in a second study in which each student received both time-limit and no-time-limit conditions. In a third study each testee received the same testing condition twice, and higher response rates were observed under the time-limit condition; response accuracy remained consistent across testing conditions. All three studies showed essentially zero correlations between response rate and response accuracy. Response latency data were also analyzed in the three studies. These data suggested the existence of different test-taking styles and strategies under time-limit and no-time-limit testing conditions. The results of these studies suggest that number-correct scores from time-limit tests are a complex function of response rate, response accuracy, test-taking style and test-taking strategy, and therefore are not likely to be as valid or as useful as number-correct scores from no-time-limit-tests. (AD A024422)

Research Report 76-3
Effects of Immediate Knowledge of Results
and Adaptive Testing on Ability Test Performance
Nancy E. Betz and David J. Weiss
June 1976

This study investigated the effects of immediate knowledge of results (KR) concerning the correctness or incorrectness of each item response on a computer-administered test of verbal ability. The effects of KR were examined on a 50-item conventional test and a stradaptive ability test and in high- and low-ability groups. The primary dependent variable was maximum likelihood ability estimates derived from the item responses. Results indicated that mean test scores for the High-Ability group receiving KR were higher than for the No-KR group on both the conventional and stradaptive tests. For Low-Ability examinees, mean scores were higher under KR conditions than under No-KR conditions on both tests, but the difference was statistically significant only for the conventional test.

However, the higher mean scores of the Low-Ability testees on the stradaptive test indicated that for low-ability examinees, adaptive testing had the same effects on test performance as did the provision of immediate KR. Knowledge of results did not have significant effects on either response latencies, response consistency on the stradaptive test, or the internal consistency reliability of the conventional test. No significant score differences were found on a 44-item post-test administered without KR, indicating that the facilitative effects of knowledge of results on test performance were confined to the test in which KR was provided. The results of the study were interpreted as indicating the potential of both immediate knowledge of results and adaptive testing procedures to increase the extent to which ability tests measure "maximum performance" levels. (AD A027147)

Research Report 76-4
Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing
Nancy E. Betz and David J. Weiss
June 1976

This study investigated the effects of providing immediate knowledge of results (KR) and adaptive testing on test anxiety and test-taking motivation. Also studied was the accuracy of student perceptions of the difficulty of adaptive and conventional tests administered with or without immediate knowledge of results. Testees were 350 college students divided into high- and low-ability groups and randomly assigned to one of four test strategies by KR conditions. The ability level of examinees was found to be related to their reported levels of motivation and to differences in reported motivation under the different testing conditions. Low-ability examinees reported significantly higher levels of motivation on the stradaptive test than on the conventional test, while the reported motivation of high-ability examinees did not differ as a function of ability level. Low-ability testees reported lower motivation with KR than without KR, while higher ability testees reported higher motivation with KR. Analysis of the anxiety data indicated that students reported significantly higher levels of anxiety on the stradaptive test than on the conventional test. The provision of KR did not result in significant differences in reported anxiety. However, highest levels of anxiety were reported by the low-ability group on the stradaptive test administered with KR. These results, in conjunction with previously reported data on effects of KR on ability test performance, were interpreted as being the result of facilitative anxiety. Students were able to perceive the relative difficulty of test items with some accuracy. However, perceptions of the relative degree of test difficulty were much more closely related to actual test score on the conventional test than on the stradaptive test. Over 90% of the students reacted favorably to the provision of immediate KR. These results suggest that adaptive testing creates a psychological environment for testing which is more equivalently motivating for examinees of all ability levels and results in a greater standardization of the test-taking environment, than does conventional testing. (AD A027170)

Research Report 77-1
Applications of Computerized Adaptive Testing
James R. McBride, James B. Sympson,
C. David Vale, Steven M. Pine, and Isaac I. Bejar
Edited by David J. Weiss
March 1977

This symposium consisted of five papers:

1. James R. McBride: A Brief Overview of Adaptive Testing
   Adaptive testing is defined, and some of its item selection and scoring strategies briefly discussed. Item response theory, or item characteristic curve theory, which is useful for the implementation of adaptive testing is briefly described. The concept of "information" in a test is introduced and discussed in the context of both adaptive and conventional tests. The advantages of adaptive testing, in terms of the nature of information it provides, are described.

2. James B. Sympson: Estimation of Latent Trait Status in Adaptive Testing Procedures
   The role of latent trait theory in measurement for criterion prediction and in criterion-referenced measurement is explicated. It is noted that latent trait models allow both normed-referenced and criterion-referenced interpretations of test performance. Using a 3-parameter logistic test model, an example of sequential estimation in a 20-item adaptive test is presented. After each item is administered, four different ability estimates (two likelihood-based and two Bayesian estimates) are calculated. Characteristics of the four estimation methods are discussed. The information available in the items selected by the adaptive test is compared with the information available from application of latent trait theory, and adaptive testing is advocated as a useful approach to human assessment.

3. C. David Vale: Adaptive Testing and the Problem of Classification
   The use of adaptive testing procedures to make ability classification decisions (i.e., cutting score decisions) is discussed. Data from computer simulations comparing conventional testing strategies with an adaptive testing strategy are presented. These data suggest that, although a conventional test is as good as an adaptive test when there is one cutting score at the middle of the distribution of ability, an adaptive test can provide better classification decisions when there is more than one cutting score. Some utility considerations are also discussed.

4. Steven M. Pine: Applications of Item Characteristic Curve Theory to the Problem of Test Bias
   It is argued that a major problem in current efforts to develop less biased tests is an over-reliance on classical test theory. Item characteristic curve (ICC) theory, which is based on individual rather than group-oriented measurement, is offered as a more appropriate measurement model. A definition of test bias based on ICC theory is presented. Using this definition, several empirical tests for bias are presented and demonstrated with real test data. Additional applications of ICC theory to the problem of test bias are also discussed.

5. Isaac I. Bejar: Applications of Adaptive Testing in Measuring Achievement and Performance
The paper reviews two relatively recent developments in psychometric theory--the assessment of partial knowledge and research in adaptive testing. It is argued that the use of non-dichotomous item formats, needed for the assessment of partial knowledge, and now made possible by the administration of achievement test items on interactive computers, should result in achievement test scores which are a more realistic and precise indication of what a student can do.
(AD A038114)

Research Report 77-2
A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items
C. David Vale and David J. Weiss
April 1977

Twenty multiple-choice vocabulary items and 20 free-response vocabulary items were administered to 660 college students. The free-response items consisted of the stem words of the multiple-choice items. Testees were asked to respond to the free-response items with synonyms. A computer algorithm was developed to transform the numerous free-responses entered by the testees into a manageable number of categories. The multiple-choice and the free-response items were then calibrated according to Bock's polychotomous logistic model. One item was discarded because of extremely poor fit with the model, and test information functions were determined from the other 19 items. Higher levels of information were obtained from the free-response items over most of the range of abilities between $\theta = -3.0$ to $\theta = +3.0$.

Research Report 77-3
Accuracy of Perceived Test-Item Difficulties
J. Stephen Prestwood and David J. Weiss
May 1977

This study investigated the accuracy with which testees perceive the difficulty of ability-test items. Two 41-item conventional tests of verbal ability were constructed for administration to testees in two ability groups. Testees in both the high- and low-ability groups responded to each multiple-choice item by choosing the correct alternative and then rating the item's difficulty relative to their levels of ability. Least-squares estimates of item difficulty, which were based on the difficulty ratings, correlated highly with proportion-correct and latent trait estimates of item difficulty based on a norming sample. Least-squares estimates of testee ability, which were based solely on the difficulty perceptions of the testees, correlated significantly with number-correct and maximum-likelihood ability scores based on the testees' conventional responses to the items. These results show that item-difficulty perceptions were highly related to the "objective" indices of item difficulty often used in test construction, and that as testee ability level increased, the items were perceived as being relatively less difficult. The relationship between a testee's ability

and his/her perception of an individual item's relative difficulty appeared to be weak. Of major importance was the finding that items which were appropriate in difficulty levels from a psychometric standpoint were perceived by the testees as being too difficult for their ability levels. The effects on testees of tailoring a test such that items are perceived as being uniformly too difficult should be investigated. (AD A041084)

Research Report 77-4
A Rapid Item-Search Procedure for Bayesian Adaptive Testing
C. David Vale and David J. Weiss
May 1977

An alternative item-selection procedure for use with Owen's Bayesian adaptive testing strategy is proposed. This procedure is, by design, faster than Owen's original procedure because it searches only part (as compared with all) of the total item pool. Item selections are, however, identical for both methods. After a conceptual development of the rapid-search procedure, the supporting mathematics are presented. In a simulated comparison with three item pools, the rapid-search procedure required as little as one-tenth the computer time as Owen's technique. (AD A041090)

Research Report 78-2
The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing
J. Stephen Prestwood and David J. Weiss
September 1978

Students were administered one of three conventional or one of three stradaptive vocabulary tests with or without knowledge of results (KR). The three tests of each type differed in difficulty, as assessed by the expected proportion of correct responses to the test items. Results indicated that the mean maximum-likelihood estimates of individuals' abilities varied as a joint function of KR-provision and test difficulty. Students receiving KR scored highest on the most-difficult test and lowest on the least-difficult test; students receiving no KR scored highest on the least-difficult test and did most poorly on the most-difficult test. Although the students perceived the differences in test difficulty, there were no effects on mean student anxiety or motivation scores attributable to difficulty alone. Regardless of test difficulty, students reacted very favorably to receiving KR, and its provision increased the mean level of reported motivation.

Research Report 79-7
The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models
Tom E. Trabin and David J. Weiss
December 1979

This study investigated a method of determining the fit of individuals to item characteristic curve (ICC) models using the person response curve (PRC). The

construction of observed PRCs is based on an individual's proportion correct on test item subsets (strata) that differ systematically in difficulty level. A method is proposed for identifying irregularities in an observed PRC by comparing it with the expected PRC predicted by the three-parameter logistic ICC model for that individual's ability level. Diagnostic potential of the PRC is discussed in terms of the degree and type of deviations of the observed PRC from the expected PRC predicted by the model.

Observed PRCs were constructed for 151 college students using vocabulary test data on 216 items of wide difficulty range. Data on students' test-taking motivation, test-taking anxiety, and perceived test difficulty were also obtained. PRCs for the students were found to be reliable and to have shapes that were primarily a function of ability level. Three-parameter logistic model expected PRCs served as good predictors of observed PRCs for over 90% of the group. As anticipated from this general overall fit of the observed data to the ICC model, there were no significant correlations between degree of non-fit and test-taking motivation, test-taking anxiety, or perceived test difficulty. Using split-pool observed PRCs, a few students were identified who deviated significantly from the expected PRC.

The results of this study suggested that three-parameter logistic expected PRCs for given ability levels were good predictors of test response profiles for the students in this sample. Significant non-fit between observed and expected PRCs would suggest the interaction of additional dimensions in the testing situation for a given individual. Recommendations are made for further research on person response curves.

Research Report 80-2
Interactive Computer Administration of a Spatial Reasoning Test
Austin T. Church and David J. Weiss
April 1980

This report describes a pilot study on the development and administration of a test using a spatial reasoning problem, the 15-puzzle. The test utilized the on-line capabilities of a real-time computer (1) to record an examinee's progress on each problem through a sequence of problem-solving "moves" and (2) to collect additional on-line data that might be of relevance to the evaluation of examinee performance (e.g., number of illegal and repeated moves, response latency trends). The examinees, 61 students in an introductory psychology class, were required to type a sequence of moves that would bring one 4 × 4 array of scrambled numbers (start configuration) into agreement with a second 4 × 4 array (goal configuration), using as few moves as possible. Data analyses emphasized the comparison of several methods of indexing problem difficulty, methods of scoring individual performance, and the relationship between response latency data, performance, and problem-solving strategy.

Subjective ratings of the perceived difficulty of replications of the 15-puzzle were obtained from a separate student sample to investigate (1) the subjective dimensions used by students in evaluating the difficulty of this problem type, (2) how accurately the actual performance difficulty of these problems could be evaluated by students, and (3) whether there were reliable individual differ-

ences in difficulty perceptions related to actual performance differences.

Results of the study suggested that four performance indices might be useful in indexing problem difficulty: (1) mean number of moves in the sample, (2) proportion of students solving the problem, (3) proportion of students solving the problem in the optimal number of moves, and (4) a Special Difficulty Index, defined as the sample mean number of moves divided by the minimum number of moves required. Four alternative methods of scoring total test performance and two methods of scoring individual problem performance were studied. The scores that took into account differential numbers of moves between the optimal and maximum number allowed were related somewhat more to performance ratings obtained from independent judges.

Examination of problem performance indices, the Special Difficulty Index, and students' perceptions of the difficulty of the test problems indicated that most of the problems were too easy for most students. However, the possibility of obtaining a more discriminating subset of problems was suggested by item-total score correlations obtained for each problem. The data suggested that better consistency might be obtained using problems of similar difficulty levels, and it was hypothesized that an adaptive test tailoring problems to the ability level of each student would increase the reliability of measurement.

Mean initial and total "move" latencies for each problem were strongly related to some of the performance indices of problem difficulty. At the level of individual performance, only total latency or problem solution time was related to problem performance. Latency data appeared to confound differences in the ability to visualize a sequence of moves and differences in students' work styles. Strong evidence for these work styles was found in student consistency of initial, average, and total response latency measures across all problems.

Perceived difficulty ratings showed reliable individual differences in the level and variability of difficulty perceptions. The data suggested that the individual differences found were related to individual differences in ability to visualize and to maintain a sequence of moves in short-term memory. It was concluded that an adequate selection of problem replications should be able to tap these differences, resulting in reliable solution performance differences.

Improvements in problem selection and design were suggested by the data in this study. Future tests of this type should consist of fewer but more difficult problems, particularly problems not permitting reactive, impulsive solutions. This type of test would seem especially appropriate for adaptive administration: (1) scores on problems tailored to the individual's ability would likely be more highly related to each other, resulting in more highly reliable total scores; (2) the motivational aspects of the tests, which seem more taxing and potentially frustrating than conventional item formats, would likely be improved, and (3) for most testees equally precise measurements could be obtained in shorter periods of time than with conventional test administration.

Research Report 80-3
Criterion-Related Validity of Adaptive Testing Strategies
Janet G. Thompson and David J. Weiss
June 1980

Criterion-related validity of two adaptive tests was compared with a conventional test in two groups of college students. Students in Group 1 (N = 101) were administered a stradaptive test and a peaked conventional test; students in Group 2 (N = 131) were administered a Bayesian adaptive test and the same peaked conventional test. All tests were computer-administered multiple-choice vocabulary tests; items were selected from the same pool, but there was no overlap of items between the adaptive and conventional tests within each group. The stradaptive test item responses were scored using four different methods (two mean difficulty scores, a Bayesian score, and maximum likelihood) with two different sets of item parameter estimates, to study the effects on criterion-related validity of scoring methods and/or item parameter estimates. Criterion variables were high school and college grade-point averages (GPA), and scores on the American College Testing Program (ACT) achievement tests.

Results indicated generally higher validities for the adaptive tests; at least one method of scoring the stradaptive tests resulted in higher correlations than the conventional test with seven of the eight criterion variables (and equal correlations for the eighth), even though the stradaptive test administered over 25% fewer items, on the average, than did the conventional test. The stradaptive test obtained a significantly higher correlation with overall college GPA (r = .27) than did the conventional test; when math GPA was partialled from overall GPA, the maximum correlation for the stradaptive test with an average length of 29.2 items was r = .51, while the 40-item conventional test correlated only .36. The data showed generally higher criterion-related validities for the mean difficulty scores on the stradaptive test in comparison to the Bayesian and maximum likelihood scores; the different item parameter estimates had no effect on validity, resulting in scores that correlated .98 with each other.

Although the mean length of the Bayesian adaptive test was 48.7 items, the median number of items (35) was less than that of the 40-item conventional test. Ability estimates from this adaptive test also correlated higher with seven of the eight criterion variables than did scores on the conventional tests, although none of the differences were statistically significant.

These data indicate that adaptive tests can achieve criterion-related validities equal to, and in some cases significantly greater than, those obtained by conventional tests while administering up to 27% fewer items, on the average. The data also suggest that latent-trait-based scoring of stradaptive tests may not be optimal with respect to criterion-related validity. Limitations of the study are discussed and suggestions are made for additional research. (AD A087595)

Research Report 80-5
An Alternate-Forms Reliability and Concurrent Validity Comparison of Bayesian Adaptive and Conventional Ability Tests
G. Gage Kingsbury and David J. Weiss
December 1980

Two 30-item alternate forms of a conventional test and a Bayesian adaptive test were administered by computer to 472 undergraduate psychology students. In addition, each student completed a 120-item paper-and-pencil test, which served as a concurrent validity criterion test, and a series of very easy questions designed to detect students who were not answering conscientiously. All test items were five-alternative multiple-choice vocabulary items. Reliability and concurrent validity of the two testing strategies were evaluated after the administration of each item for each of the tests, so that trends indicating differences in the testing strategies as a function of test length could be detected. For each test, additional analyses were conducted to determine whether the two forms of the test were operationally alternate forms.

Results of the analysis of alternate-forms correspondence indicated that for all test lengths greater than 10 items, each of the alternate forms for the two test types resulted in fairly constant mean ability level estimates. When the scoring procedure was equated, the mean ability levels estimated from the two forms of the conventional test differed to a greater extent than those estimated from the two forms of the Bayesian adaptive test.

The alternate-forms reliability analysis indicated that the two forms of the Bayesian test resulted in more reliable scores than the two forms of the conventional test for all test lengths greater than two items. This result was observed when the conventional test was scored either by the Bayesian or proportion-correct method.

The concurrent validity analysis showed that the conventional test produced ability level estimates that correlated more highly with the criterion test scores than did the Bayesian test for all lengths greater than four items. This result was observed for both scoring procedures used with the conventional test.

Limitations of the study, and the conclusions that may be drawn from it, are discussed. These limitations, which may have affected the results of this study, included possible differences in the alternate forms used within the two testing strategies, the relatively small calibration samples used to estimate the ICC parameters for the items used in the study, and method variance in the conventional tests. (AD A094477)

Research Report 81-2
Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing
Marilyn F. Johnson, David J. Weiss, and J. Stephen Prestwood
February 1981

The study investigated the joint effects of knowledge of results (KR or no-KR), pacing of item presentation (computer or self-pacing), and type of testing

strategy (50-item peaked conventional, variable-length stradaptive, or 50-item fixed-length stradaptive test) on ability test performance, test item response latency, information, and psychological reactions to testing. The psychological reactions to testing were obtained from Likert-type items that assessed test-taking anxiety, motivation, perception of difficulty, and reactions to knowledge of results. Data were obtained from 447 college students randomly assigned to one of the 12 experimental conditions.

The results indicated that there were no effects on ability estimates due to knowledge of results, testing strategy, or pacing of item presentation. Although average latencies were greater on the stradaptive tests than on the conventional test, the overall testing time was not substantially longer on the adaptive tests and may have been a function of differences in test difficulty. Analysis of information values indicated higher levels of information on the stradaptive tests than on the conventional test. There was no statistically significant main effect for any of the three experimental conditions when test anxiety or test-taking motivation were the dependent variables, although there were some significant interaction effects.

These results indicate that testing conditions may interact in a complex way to determine psychological reactions to the testing environment. The interactions do suggest, however, a somewhat consistent standardizing effect of KR on test anxiety and test-taking motivation. This standardizing effect of KR showed that approximately equal levels of motivation and anxiety were reported under the various testing conditions when KR was provided, but that mean levels of these variables were substantially different when KR was not provided. Consistent with theoretical expectations, the conventional test was perceived as being either too easy or too difficult, whereas the adaptive tests were perceived more often as being of appropriate difficulty.

The results concerning the effects of KR on test performance, motivation, and anxiety found in this study were contrary to earlier reported findings; and differences in the studies are delineated. Recommendations are made concerning the control of specific testing conditions, such as difficulty of the test and ability level of the examinee population, as well as suggestions for the further analysis of the standardizing effect of KR.

Research Report 83-1
Reliability and Validity of Adaptive and Conventional Tests
In a Military Recruit Population
John T. Martin, James R. McBride, and David J. Weiss
January 1983

A conventional verbal ability test and a Bayesian adaptive verbal ability test were compared using a variety of psychometric criteria. Tests were administered to 550 Marine recruits, half of whom received two 30-item alternate forms of a conventional test and half of whom received two 30-item alternate forms of a Bayesian adaptive test. Both types of tests were computer administered and were followed by a 50-item conventional verbal ability criterion test.

The alternate forms of the adaptive test resulted in scores that were much more

similar in means and variances than were the conventional tests for which most means and variances for various test lengths were significantly different. Adaptive testing resulted in significantly higher alternate forms reliability correlations for all test lengths through 19 items; reliability of a 9-item adaptive test was equal to that of a 17-item conventional test. Validity correlations were higher for the adaptive procedure for all test lengths. Validity of an 11-item adaptive test was equal to that of a 27-item conventional test, in spite of lower discriminating items being used, on the average, by the adaptive tests in comparison to the conventional test. Very few of the recruits had difficulty in responding to the computer-administered instructions on use of the testing terminals. Analysis showed some differences in test duration between the two testing strategies; where they occurred, they were explained by the ability level of the examinees, i.e., higher ability examinees who were administered adaptive tests received more difficult items and therefore had significantly longer testing times. Combined with reduced test length for the adaptive test to obtain similar reliabilities and validities to the conventional test, however, the slight increases observed in adaptive testing time were negligible.

The data support the feasibility of adaptive testing with military recruit populations and support theoretical predictions of the psychometric superiority of adaptive tests in comparison with number-correct scored conventional tests. (AD A129324)

Distribution List

Navy

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

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

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

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

1 Lt. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32509

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

1 Dr. Robert Carroll
NAVOP 115
Washington , DC 20370

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

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

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

1 Dr. Doug Davis
CNET
Pensacola, FL

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

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

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

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

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

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

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

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

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

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

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

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

1 Bill Nordbrock
1032 Fairlawn Ave.
Libertyville, IL 60048

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Marine Corps

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

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

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

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

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

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

Army

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Air Force

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1 Major John Welsh
AFMPC
Randolph AFB, TX

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

Department of Defense

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

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

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

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

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

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

Civilian Agencies

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

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

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

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

Private Sector

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1 Dr. Douglas H. Jones
Advanced Statistical Technologies
Corporation
10 Trafalgar Court

Lawrenceville, NJ 08148

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

PREVIOUS PUBLICATIONS (CONTINUED)

78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-4. 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 1979 Computerized Adaptive Testing Conference. September 1980

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

Research Reports

83-3. Effect of Examinee Certainty on Probabilistic Test Scores and a Comparison of Scoring Methods for Probabilistic Responses. July 1983.

83-2. Bias and Information of Bayesian Adaptive Testing. March 1983.

83-1. Reliability and Validity of Adaptive and Conventional Tests in a Military Recruit Population. January 1983.

81-5. Dimensionality of Measured Achievement Over Time. December 1981.

81-4. Factors Influencing the Psychometric Characteristics of an Adaptive Testing Strategy for Test Batteries. November 1981.

81-3. A Validity Comparison of Adaptive and Conventional Strategies for Mastery Testing. September 1981.

Final Report: Computerized Adaptive Ability Testing. April 1981.

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.

-continued inside-

END

FILMED

9-84

DTIC