DOCUMENT RESUME

ED 209 354		TM 810 916
AUTHOR	Reckase, Mark D.	
TITLE	Procedures for Criterion-Referenced Tailored Testing. Final Report, 1977-1981.	
INSTITUTION	Missouri Univ., Columbia. Dept. of Educational Psychology.	
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.	
PUB DATE	Aug 81	
CONTRACT	N00014-77-C-0097	
NOTE	26p.	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	Achievement Tests; *Criterion Referenced Tests; Decision Making; *Factor Analysis; *Latent Trait Theory; Maximum Likelihood Statistics; Models; *Test Construction	
IDENTIFIERS	ANCILLES Estimation Procedures; Item Calibration; LOGIST Computer Program; *Tailored Testing; Unidimensional Scaling	

ABSTRACT

This report summarizes the research findings of a four year contract investigating the applicability of item response theory and tailored testing to criterion-referenced measurement. Six major areas were studied on the project: (1) techniques for forming unidimensional item sets; (2) techniques for calibrating items; (3) item parameter linking procedures; (4) comparisons of latent trait models; (5) tailored testing procedures; and (6) decision making procedures. Results showed that factor analytic procedures were best at forming unidimensional item pools, the LOGIST calibration program performed slightly better than the ANCILLES program for item calibration, the maximum likelihood procedure using the LOGIST program generally gave the best linking, the three-parameter logistic model was preferred to the one-parameter model for tailored testing applications, the maximum likelihood based tailored testing procedure was slightly preferred to the Owen's Bayesian based procedure, and the use of the sequential probability ratio test with tailored testing resulted in substantial savings in test length. Overall, tailored testing was shown to be feasible for achievement testing applications. More detailed results are described in the papers and reports listed in this report. (Author/GK)

Final Report: Procedures for Criterion Referenced Tailored Testing

Mark D. Reckase

Final Report
August 1981

Tailored Testing Research Laboratory
Educational Psychology Department
University of Missouri
Columbia, MO 65211

Prepared under contract No. N00014-77-C-0097, NR150-395
with the Personnel and Training Research Programs
Psychological Science Division

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	
2. GOVT ACCESSION NO.	
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	Final Report: Procedures for criterion referenced tailored testing
5. TYPE OF REPORT & PERIOD COVERED	Final Report 1977-1981
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	Mark D. Reckase
8. CONTRACT OR GRANT NUMBER(s)	N00014-77-C-0097
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Department of Educational Psychology University of Missouri Columbia, MO 65211
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	P.E.: 61153N Proj.: RR042-04 T.A.: 042-04-01 W.U.: NR150-395
11. CONTROLLING OFFICE NAME AND ADDRESS	Personnel and Training Research Programs Office of Naval Research
12. REPORT DATE	August 1981
13. NUMBER OF PAGES	11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION DOWNGRADING SCHEDULE	

16. DISTRIBUTION STATEMENT (of this Report)

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

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

18. SUPPLEMENTARY NOTES

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

Testing	Item Calibration
Achievement Testing	Linking
Latent Trait Models	Sequential Probability Ratio Test
Tailored Testing	Factor Analysis

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

This report summarizes the research findings of a four year contract investigating the applicability of item response theory and tailored testing to criterion-referenced measurement. Six major areas were studied on the project. These included: (a) techniques for forming unidimensional item sets, (b) techniques for calibrating items, (c) item parameter linking procedures, (d) comparisons of latent trait models, (e) tailored testing procedures, and (f) decision making procedures. The results showed that

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) factor analytic procedures were best at forming unidimensional item pools, the LOGIST calibration program performed slightly better than the ANCILLES program for item calibration, the maximum likelihood procedure using the LOGIST program generally gave the best linking, the three-parameter logistic model was preferred to the one-parameter model for tailored testing applications, the maximum likelihood based tailored testing procedure was slightly preferred to the Owen's Bayesian based procedure, and the use of the sequential probability ratio test with tailored testing resulted in substantial savings in test length. Overall, tailored testing was shown to be feasible for achievement testing applications. More detailed results are described in the papers and reports listed in this report.

CONTENTS

FINAL REPORT:

PROCEDURES FOR CRITERION-REFERENCED TAILORED TESTING

The purpose of this contract has been to investigate the applicability of item response theory (IRT) and tailored testing to criterion-referenced measurement. Since criterion-referenced measurement involves creating an item domain, setting criterion cutoffs, and making a decision as to the location of an examinee relative to the cutoff, the applicability of item response theory and tailored testing had to be evaluated for each of these components.

The investigation of the first component, creating an item domain, involved evaluating procedures for forming unidimensional item sets, procedures for item calibration and procedures for linking calibrations together to form large item pools. This was done since IRT requires an assumption of unidimensionality, and large item pools are required for tailored testing. The setting of criterion cutoffs and the decision making aspect of criterion-referenced testing required the investigation of the IRT/tailored testing approach to achievement testing and to decision making. Therefore, the various IRT and tailored testing models were evaluated for achievement testing applications and a decision making procedure was developed for tailored testing applications. Each of these components will now be described in detail and research results will be summarized.

Formation of Unidimensional Item Sets

Because of the assumption that the items used in an IRT based procedure can be described in a unidimensional latent space, it was considered an important component of this project to evaluate procedures for selecting test items to meet this assumption. Therefore, a study was planned to evaluate the ability of various procedures to determine the dimensionality of a set of test items and to sort items into sets measuring a single dimension. The procedures evaluated included factor analysis, nonmetric multidimensional scaling, cluster analysis, and latent trait theory analysis. These procedures were applied to simulated and real test data of varying factorial complexity. In all cases, guessing was present in the data since multiple choice items were assumed for the tailored testing application.

The results of this study are reported in:

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

and in:

Reckase, M. D. Guessing and dimensionality: The search for a unidimensional latent space. Paper presented at the meeting of the American Educational Research Association, Los Angeles, April 1981.

Reckase, M. D. The effect of guessing in dichotomously scored items on the operation of multivariate data reduction techniques. Paper presented at the meeting of the Psychometric Society, Iowa City, IA, May 1980.

The results indicated that factor analytic and nonmetric multidimensional scaling techniques could be used to sort items into unidimensional sets and that cluster analysis and latent trait analysis were generally not appropriate. Of the factor analytic techniques evaluated, principal factor analysis of phi-coefficients was found to give the best information for determining the dimensionality of a set of items. Nonmetric multidimensional scaling was found to work well with some similarity coefficients while giving fairly meaningless results with others. The MDSCAL program (Kruskal, 1964) used gave best results with the Yule's Y coefficient, phi-coefficient and tetrachoric correlations.

The cluster analysis procedure was found to be inappropriate because of difficulty in determining how many clusters should be present in the data. Both hierarchial and complete link procedures were evaluated. The latent trait procedures worked fairly well. but were too cumbersome for general use. The procedure involved successive applications of LOGIST (Wood, Wingersky, and Lord, 1976) to a set of test items, deleting the items with low a-values after each run. Unidimensional subsets were usually formed in this way, but as many as ten program runs were required for each item set. This was clearly impractical, especially considering the cost of the program runs.

Item Calibration

Once it had been determined that a set of items met the assumption of a unidimensional latent space, the items needed to be calibrated according to one of the IRT models. That is, the parameters of the model needed to be estimated using one of the available computer programs for that purpose. There are several IRT models that could be used for tailored testing applications and each of these models has several calibration programs for use in estimating its parameters. One of the initial tasks of this contract was to compare the various models available and to evaluate their calibration programs.

The results of a review of the literature and a comparison of item calibrations models are presented in:

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

and

McKinley, R. L. & Reckase, M. D. A comparison of the ANCILLES and LOGIST parameter estimation procedures for the three-parameter logistic model using goodness of fit as a criterion (Research Report 80-2). Columbia, MO: University of Missouri, December 1980.

Other papers related to this topic are:

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Measurement, 1979, 4(3), 207-230.

McKinley, R. L. & Reckase, M. D. The fit of ICC's based on two different three-parameter logistic model parameter estimation procedures. Paper presented at the meeting of the Psychometric Society, Chapel Hill, N.C., May 1981.

Reckase, M. D. The validity of latent trait models through the analysis of fit and invariance. Paper presented at the meeting of the American Educational Research Association, Los Angeles, April 1981.

Reckase, M. D. A comparison of the one- and three-parameter logistic models for item calibration. Paper presented at the meeting of the American Educational Research Association, Toronto, March 1978.

Reckase, M. D. Univariate latent trait models applied to multivariate measures. Paper presented at the meeting of the Psychometric Society, Chapel Hill, N.C., May 1977.

The results of the research on calibration techniques can be divided into two parts: the comparison of calibration models, and the comparison of calibration programs for a single model. The comparison of calibration models concentrated on the one- and three-parameter logistic models. The results of the research on the calibration models showed that the three-parameter model fit empirical data better than the one-parameter model, but that the sample size required to use the three-parameter model was substantially larger than for the one-parameter model. Lack of fit was most prominent for the one-parameter model when guessing was a factor in item response.

The models were also found to yield ability estimates that measured different constellations of ability when items tapping several dimensions were used in a test. The one-parameter logistic based ability estimates were related to the sum of the components present in a test, while the three-parameter logistic based ability estimates were found to be mainly related to the single largest component in a test. This difference in ability estimates is due to differences in the weighting of the item responses for the two models. Unit weights are used for the one-parameter model, while the items are weighted by the item discrimination parameter estimates for the three-parameter model. Despite these differences, the ability estimates obtained from the models were found to be highly correlated for many tests composed of a fixed set of items. The controlling factor seemed to be the magnitude of the first principal component of the test. When the item calibration results from multiple choice tests were to be used for tailored testing purposes, the three-parameter logistic model was found to be superior because of the better fit to empirical data. The ability to approximate guessing effects was found to be especially important because the error induced in the one-parameter item parameter estimates by this factor.

Among the numerous available item calibration procedures, the ANCILLES (Urry, 1978) and LOGIST (Wood, Wingersky & Lord, 1976) procedures were selected for comparison on this project because of their wide useage by the testing community. The results of the research showed that the ICC estimates from the LOGIST program fit the empirical item data slightly better than those from the ANCILLES program. For this reason, the LOGIST program was suggested for item calibration for use with tailored testing procedures.

In addition to comparing existing calibration models and programs, a new estimation procedure was developed on the project by Robert Tsutakawa and Steve Rigdon. This new procedure is described in detail in the report:

Rigdon, S. E. & Tsutakawa, R. K. Estimation in latent trait models (Research Report 80-1). Columbia, MO: University of Missouri, May 1981.

The investigation focused on the estimation of ability and item parameters for a class of binary response models, including the commonly used logistic and probit models. Estimation procedures were examined for variations of these models depending on whether only the ability parameters or both ability and item parameters are assumed random with prior distributions with fixed but unknown hyperparameters.

When the item parameters are fixed and the ability parameters are random, the EM algorithm can be readily adapted. Estimates of ability parameters are easily found, even in situations where maximum likelihood estimates do not exist. Simulation studies have shown that these estimates are more efficient than maximum likelihood estimates in terms of the mean square error criterion. The EM algorithm can be modified to estimate the item parameters by conditioning on the expected ability parameters. This revised method is computationally much cheaper while performing as well as the straight EM algorithm. Although most of the numerical work has been restricted to the one-parameter logistic (Rasch) model, with a normal prior, the method extends to multi-parameter models. Computer programs for the two-parameter (for ability and guessing) logistic model are now being developed.

When both item and ability parameters are considered random, the EM algorithm applies in principle but cannot be easily implemented since the random variables do not have distributions belonging to exponential families. The algorithm was modified by alternately estimating the ability and item parameters while holding one set fixed at its posterior expectation. Simulated results assuming normal priors indicated that the resulting estimators do not perform as well as the maximum likelihood estimator. This discrepancy disappears, however, when the prior distribution of the difficulty parameter was assumed uniform.

The extent to which the models used here are applicable in practice remains to be seen. Some preliminary work was done on goodness of fit tests. Though it appears that the classical chi-square methods apply when ability parameters are from a common prior distribution, the amount of computation needed for even a moderate number of items may be prohibitive. A more feasible approach may be through the logarithmic penalty function mentioned by Mosteller and Wallace (1964) in their book on the Federalist Papers and examined in more detail by Efron (1978).

Two other areas included in the original research objectives were comparing item response curves and designing sequential methods for mental testing. Work was limited since the procedures depend on the results from the EM algorithm study.

Item Parameter Linking

Once item parameter estimates had been obtained from a series of group tests, they needed to be placed on the same scale (linked) so all of the items could be used in the tailored testing item pool. Numerous procedures have been developed for this linking task. A natural extension of the research on models and calibration procedures was to evaluate linking procedures to determine which gave the most accurate parameter estimates for use with tailored testing. The results of the evaluation were reported in the following report:

McKinley, R. L. & Reckase, M. D. A comparison of procedures for constructing large item pools (Research Report 81-3). Columbia, MO: University of Missouri, August 1981.

Some of the results of this research were also reported in:

Reckase, M. D. Item pool construction for use with latent trait models. Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

The basic design for this part of the research effort was to sample a series of short tests from a long test, link the calibrations of the short tests, and then compare the linked parameter estimates to those obtained from the long test. This procedure was used to evaluate linking procedures for both the one-parameter and three-parameter models and to determine the necessary sample size and number of common items between tests for linking to be performed. The MAX calibration program (Wright & Panchapakesan, 1969) was used for the one-parameter model and the ANCILLES and LOGIST programs were used for the three-parameter model.

The results of the research showed that far fewer cases were required to link the single parameter of the one-parameter model than were needed for the parameters of the three-parameter model. Approximately 2000 cases seemed to be needed in the latter case. Of the four linking procedures evaluated for the three-parameter logistic model, maximum likelihood linking using the LOGIST program gave the best results overall when a 2000 sample was used. Fifteen items in common between test forms were sufficient for adequate linking. Future research should address the quality of items that should be in common between tests.

Latent Trait Model

Once an item pool has been produced using the calibration and linking methods described above, the actual tailored testing process can begin. To define that process, one of the many latent trait models must be selected as a basis for the tailored testing procedure. Of the many models available, the one- and three-parameter logistic models were evaluated for use in this project. A series of three tailored testing studies were run to determine which of these two models gave the most accurate ability estimates in a realistic testing setting. The following reports describe these studies in detail.

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

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

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

Other papers written on this topic were:

McKinley, R. L. & Reckase M. D. Computer application to ability testing. AEDS Journal, 1980, 13(3), 1_3-203.

Reckase, M. D. Procedures for computerized testing. Behavior Research Methods and Instrumentation, 1977, 9(2), 148-152.

English, R. A., Reckase, M. D. & Patience, W. M. Application of tailored testing to achievement measurement. Behavior Research Methods and Instrumentation, 1977, 9(2), 158-161.

Reckase, M. D. Tailored testing, measurement problems, and latent trait theory. Paper presented at the meeting of the National Council on Measurement in Education, Los Angeles, April 1981.

Patience, W. M. & Reckase, M. D. Self-paced versus paced evaluation utilizing computerized tailored testing. Paper presented at the meeting of the National Council on Measurement in Education, Toronto, March 1978.

Reckase, M. D. Computerized achievement testing using the simple logistic model. Paper presented at the meeting of the American Educational Research Association, New York, April 1977.

The one- and three-parameter logistic based tailored testing procedures used in these studies were both based on maximum information item selection and maximum likelihood ability estimation. The criteria for evaluation of the ability estimates obtained from the procedures were the information function and reliability coefficients obtained when the procedures were applied in a realistic setting. Both vocabulary and achievement items were used in the evaluation. The populations used for the studies were upper level college students.

The overall results obtained from the series of studies showed that a tailored testing procedure based on the three-parameter logistic model gave both higher information values and higher reliability coefficients than the one-parameter model. The predictive validity of the ability estimate using scores on classroom achievement tests as a criterion was found to be about equal for the two models. Twenty item tailored tests were found to give about equivalent reliability to 50 item traditional tests on the same material.

An important finding of the live testing research on tailored testing was the determination of the sensitivity of the procedures to the accuracy of the item calibration information. When item parameter estimates were poor, the procedures gave meaningless results, regardless of the quality of the test items. Inaccurate parameter estimates also made the information function meaningless. High information values were sometimes obtained for tests with low reliabilities. These results point out the critical importance of item calibration and linking.

In addition to evaluating the quality of ability estimates using the two procedures, research was done to determine the best way to operate the tailored testing procedure. This research involved determining the composition of the item pool, the appropriate place in the pool to start administering items, and the item selection procedure to use before ability estimates were available. The results of this research are reported in the following reports and papers.

Patience, W. M. & Reckase, M. D. Effects of program parameters and item pool characteristics on the bias of a three-parameter tailored testing procedure. Paper presented at the meeting of the National Council on Measurement in Education, Boston, April 1980.

Patience, W. M. & Reckase, M. D. Operational characteristics of a one-parameter tailored testing procedure (Research Report 79-2). Columbia, MO: University of Missouri, October 1979.

Patience, W. M. & Reckase, M. D. Operational characteristics of a Rasch model tailored testing procedure when program parameter and item pool attributes are varied. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, April 1979.

The methodology used for this research was to simulate the testing process for a hypothetical examinee of known ability and determine the mean and standard deviation of the obtained ability estimates. The procedure was considered acceptable if the resulting estimates were unbiased and had a small variance. An analytic procedure that traced all possible paths through the item pool for a person of known ability was also used to determine the statistical bias and variance of estimates.

The results of this research indicated that the characteristics of the item pool are important in determining the quality of the ability estimates. The item pool should have a rectangular distribution of item difficulties, and a uniform level of item discrimination. Items with low discrimination parameter estimates are not selected by the tailored testing procedure so they should not be included in determining the size of the active item pool. It was also found to be important that the difficulty scale be uniformly covered by items. If gaps in the coverage were present, regions of the ability scale would be poorly estimated.

Other recommendations can be made based on this research. First, the initial ability estimate used to start the testing session should be one that selects a first item of about median difficulty since it is important that enough items are present both above and below the initial ability to give good estimation. Also, when using a maximum likelihood ability estimation procedure, the stepsize used before an estimate is obtained should be approximately .7 for the one-parameter procedure and .3 for the three-parameter procedure. Otherwise, the examinee's ability estimate may move out of the range where items are present before a good ability estimate can be obtained.

The recommendations given above should only be considered as rough guidelines because of the complex interaction of the variables controlling the tailored testing situation. The best recourse is to simulate the operation of about 50 examinees at numerous points along the ability scale using the actual item parameters from the item pool to be used. This procedure can be used to determine the accuracy of ability estimates that can be obtained. The controlling parameters of the tailored testing program can then be fine tuned to the item pool.

Tailored Testing Procedure

Based upon research reported up to this point, the three-parameter logistic model is a clear choice over the one-parameter logistic model for tailored testing applications. However, there are two commonly used procedures for applying the three-parameter model to tailored testing, Owen's Bayesian procedure (Owen, 1975) and the maximum likelihood procedure, and little work has been done to directly compare the two. A live testing study comparing these two procedures was conducted on this project to obtained information relevant to choosing between them. The results of the study are given in the following report and paper:

McKinley, R. L. & Reckase, M. D. A comparison of a Bayesian and a maximum likelihood tailored testing procedure (Research Report 81-2). Columbia, MO: University of Missouri, August 1981.

Rosso, M. A. & Reckase, M. D. A comparison of a maximum likelihood and a Bayesian ability estimation procedure for tailored testing. Paper presented at the meeting of the National Council on Measurement in Education, Los Angeles, April 1981.

This study compared the reliability coefficients, information functions and ability estimates for achievement tests administered using either Owen's Bayesian procedure or a maximum likelihood tailored testing procedure. The Bayesian procedure selected items to minimize the posterior variance of the ability estimates and estimated ability as the mean of the posterior distribution of ability. The maximum likelihood procedure selected items to maximize the information function at the most recent ability estimate and estimated ability using an empirical maximum likelihood approach.

The results of the study showed that the two procedures had approximately equal reliabilities and information functions. However, a definite regression effect was found to be a result of the Bayesian prior. In this study, the prior was assumed to be normal with a mean near the median difficulty of the item pool and a variance of 1.0. Since the prior mean was somewhat lower than the ability of the group tested, the ability estimates were artificially kept lower than the maximum likelihood estimates. Because of this effect, the maximum likelihood procedure was recommended if accurate prior information were not available.

Decision Making Procedure

The final project undertaken on this contract was to investigate decision making procedures for use with tailored testing. The most convenient procedure found for use with tailored testing was based on Wald's (1947) sequential probability ratio test. Research was done on the contract to determine the usefulness of such a procedure. The results of the research effort are presented in the following reports and papers.

Reckase, M. D. The use of the sequential probability ratio test in making grade classifications in conjunction with tailored testing (Research Report 81-4). Columbia, MO: University of Missouri, August 1981.

Reckase, M. D. Some decision procedures for use with tailored testing. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis, University of Minnesota, 1980.

Reckase, M. D. An application of tailored testing and sequential analysis to classification problems. Paper presented at the meeting of the American Educational Research Association, Boston, April 1980.

Reckase, M. D. A generalization of sequential analysis to decision making with tailored testing. Paper presented at the meeting of the Military Testing Association, Oklahoma City, Ok: November 1978.

The SPRT procedure was investigated for use with tailored testing using both simulation and live testing techniques. Use of the SPRT with both the one-parameter and three-parameter models was studied. The results showed that a substantial reduction of test length could be attained through the use of the SPRT without loss of decision accuracy. As with previous work, the three-parameter procedure was found to yield better results than the one-parameter procedure. The detrimental effects of guessing on the operation of the one-parameter logistic based tailored testing procedure were an especially important factor in the results.

Summary and Conclusion

This research project studied many facets of the application of IRT and tailored testing to achievement measurement. Included were studies of techniques for sorting items into unidimensional item sets, calibrating test items, linking item calibrations, estimating ability, selecting items for tailored testing, and making decisions using tailored testing. Overall, the results were fairly positive. Unidimensional item sets can be formed using the principal factor technique on phi coefficients and the items can be calibrated for use with tailored testing using the LOGIST program if a sufficient sample of individuals is available. The calibration of separate tests can be linked to produce a large item pool if at least 15 items are in common between the tests and a sample of performance for at least 2000 individuals is available for each test. The maximum likelihood procedure using the LOGIST program is recommended for the linking. The three-parameter logistic model has been shown to give an adequate theoretical basis for tailored testing, even for achievement testing which does not quite meet the assumptions of the model. A tailored testing model based on maximum information item selection and maximum likelihood ability estimation is recommended for use, with an expected result of reducing the number of items required to obtain a test reliability equal to that of a traditional test more than twice as long. Accurate decision making has been shown to be possible u ng the sequential probability ratio test with tailored testing with a substantial reduction in the number of test items administered and tight control of errors of classification.

With all of these positive results, there are still many areas in which the user of tailored testing must exercise extreme caution. Unlike traditional paper and pencil testing, tailored testing is critically dependent on the quality of the calibration and linking of the item pool. If the item parameters are poorly estimated, the item selection procedure and ability estimation procedure will be operating on meaningless numbers and will tend to give meaningless results. The situation is equivalent to determining the length of a line with a ruler that has its units marked off in the wrong places. Trusting the calibration too much can give test results that look good (i.e., have high information functions), when the reliability of the scores is in fact very low. As a suit, tailored tests should still be evaluated for quality using proc independent of the item parameters, such as test-retest reliability.

The use of the three-parameter logistic model as a basis for ability estimation causes some subtle problems in test score interpretation. Using this model causes item responses to be weighted by the discrimination parameter estimates when computing an estimate of ability. This ives high weight to those items measuring the major component of a test and very low weight to items not measuring that component. The result is an ability estimate measuring a trait that is more unidimensional than is obtained from a number correct score. This difference must be taken into account when making use of tailored testing.

The use of tailored testing for measurement is similar in many ways to the use of computers for computation. The techniques give high power and efficiency based on high technology. But a price is paid for the advantages. The price is a greater sensitivity to the input to the procedure and a greater dependence on complicated hardware. The results of this research contract definitely show that tailored testing can be applied to achievement measurement with many advantages. However, the technique cannot be applied carelessly and still achieve those advantages.

REFERENCES

Efron, B. Regression and ANOVA with zero-one data: Measures of residual variation. Journal of the American Statistical Association, 1978, 73(361), 113-121.

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129.

Mosteller, F. and Wallace, D. Inference and disputed authorship: The Federalists. Palo Alto: Addison-Wesley, Inc., 1964.

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

Urry, V. W. ANCILLES: Item parameter estimation program with normal ogive and logistic three-parameter model options. Washington, D. C.: U. S. Civil Service Commission, Personnel Research and Development Center, 1978.

Wald, A. Sequential analysis. New York: Wiley, 1947.

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

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

Navy

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

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

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

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

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

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

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

Navy

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

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 Co
Newport News, VA 23607

1 Dr. James McBride
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

Navy

1 Library, Code P201L
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 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Selection and Training Research Division
Human Performance Sciences Dept.
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

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

Navy

1 Dr. Worth Scanland, Director
Research, Development, Test & Evaluation
N-5
Naval Education and Training Command
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 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

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

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

Army

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

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

1 Dr. Dexter Fletcher
U.S. Army Research Institute
5001 Eisenhower Avenue
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. JAMES L. RANEY
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

Army

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

1 Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
FT Benjamin Harrison, IN 46256

1 Dr. Frederick Steinheiser
Dept. of Navy
Chief of Naval Operations
OP-113
Washington, DC 20350

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

Air Force

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

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

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

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

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

Marines

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

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

1 Major Michael L. Patrow, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

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

CoastGuard

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

Other DoD

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

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

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

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

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

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

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

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

Non Govt

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

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

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

1 Capt. J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario KOK 1BO

1 CDR Robert J. Biersner
Program Manager
Human Performance
Navy Medical R&D Command
Bethesda, MD 20014

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

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

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

1 Col Ray Bowles
800 N. Quincy St.
Room 804
Arlington, VA 22217

Non Govt

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

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

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

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

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

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

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

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

Non Govt

1 Dr. Fritz Drasgow
Yale School of Organization and Manageme
Yale University
Box 1A
New Haven, CT 06520

1 Dr. Mavin D. Dunnette
Personnel Decisions Research Institute
2415 Foshay Tower
821 Marguette Avenue
Mineapolis, MN 55402

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

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

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

Non Govt

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

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

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

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

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

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

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

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

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

Non Govt

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

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

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

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

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

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

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

Non Govt

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

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

1 Professor Jason Millman
Department of Education
Stone Hall
Cornell University
Ithaca, NY 14853

1 Bill Nordbrock
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

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

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

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

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

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

Non Govt

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

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

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

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

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

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

Non Govt

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

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

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

Non Govt

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

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

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

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

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

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

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