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 show 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 reference tailored testing

Mark D. Keck

department of Educational Psychology
University of Missouri
Columbia, MO 65211

Personnel and Training Research Programs
Office of Naval Research

August 1981

11

12

13

14

15

16

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

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...
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.
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:

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:

and


Other papers related to this topic are:


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 usage 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 assured 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:


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

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.


Other papers written on this topic were:


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.


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:


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 his 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 in 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. 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 using 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 result, tailored tests should still be evaluated for quality using procedures 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 gives 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


<table>
<thead>
<tr>
<th>Name</th>
<th>Title/Office</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Jack R. Borsting</td>
<td>Provost &amp; Academic Dean</td>
<td>U.S. Naval Postgraduate School</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monterey, CA 93940</td>
</tr>
<tr>
<td>Dr. Robert Breaux</td>
<td></td>
<td>Code N-711 NAVTRAEOUPCEN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Orlando, FL 32813</td>
</tr>
<tr>
<td>Chief of Naval Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp; Training Liaison Office</td>
<td>Air Force Human Resource Laboratory Flying Training Division</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WILLIAMS AFB, AZ 85224</td>
</tr>
<tr>
<td>CDR Mike Curran</td>
<td>Office Naval Research</td>
<td>800 N. Quincy St.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Code 270 Arlington, VA 22217</td>
</tr>
<tr>
<td>Dr. Richard Elster</td>
<td>Department of Administrative Sciences</td>
<td>Naval Postgraduate School</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monterey, CA 93940</td>
</tr>
<tr>
<td>DR. PAT FEDERICO</td>
<td>Navy Personnel R&amp;D Center</td>
<td>SAN DIEGO, CA 92152</td>
</tr>
<tr>
<td>Mr. Paul Foley</td>
<td></td>
<td>Navy Personnel R&amp;D Center</td>
</tr>
<tr>
<td></td>
<td></td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Dr. John Ford</td>
<td>Navy Personnel R&amp;D Center</td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Dr. Henry M. Halff</td>
<td>Department of Psychology, C-009</td>
<td>University of California at San Diego</td>
</tr>
<tr>
<td></td>
<td></td>
<td>La Jolla, CA 92093</td>
</tr>
<tr>
<td>Dr. Patrick R. Harrison</td>
<td>Psychology Course Director</td>
<td>LEADERSHIP &amp; LAW DEPT. (7b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIV. OF PROFESSIONAL DEVELOPMENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U.S. NAVAL ACADEMY</td>
</tr>
<tr>
<td>CDR Charles W. Hutchins</td>
<td>Naval Air Systems Command HQ</td>
<td>AIR-340F Navy Department</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Washington, DC 20361</td>
</tr>
<tr>
<td>Dr. Norman J. Kerr</td>
<td>Chief of Naval Technical Training</td>
<td>Naval Air Station Memphis (75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Millington, TN 38054</td>
</tr>
<tr>
<td>Dr. William L. Maloy</td>
<td>Principal Civilian Advisor for Education and Training</td>
<td>Naval Training Command, Code 00A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pensacola, FL 32508</td>
</tr>
<tr>
<td>Dr. Kneale Marshall</td>
<td>Scientific Advisor to DCNO(MPT)</td>
<td>OPO1T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Washington DC 20370</td>
</tr>
<tr>
<td>CAPT Richard L. Martin,</td>
<td>Prospective Commanding Officer</td>
<td>USS Carl Vinson (CVN-70)</td>
</tr>
<tr>
<td>USN</td>
<td></td>
<td>Newport News Shipbuilding and Drydock Co</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Newport News, VA 23607</td>
</tr>
<tr>
<td>Dr. James McBride</td>
<td>Navy Personnel R&amp;D Center</td>
<td>San Diego, CA 92152</td>
</tr>
<tr>
<td>Ted M. I. Yellen</td>
<td>Technical Information Office, Code 201</td>
<td>NAVY PERSONNEL R&amp;D CENTER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAN DIEGO, CA 92152</td>
</tr>
</tbody>
</table>
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 Street
Arlington, VA 22217

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

1 Psychologist
ONR Branch Office
103C 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 Laboratory
Pensacola, FL 32508

1 Dr. Bernard Rimland (O3B)
Navy Personnel R&D Center
San Diego, CA 92152

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 Wister
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
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 Measurement 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
Coast Guard

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 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 1B0

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

1 Dr. Manuela 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
<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dr. Robert Brennan</td>
<td>American College Testing Programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. O. Box 168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iowa City, IA 52240</td>
</tr>
<tr>
<td>1</td>
<td>Dr. C. Victor Bunderson</td>
<td>WICAT INC.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNIVERSITY PLAZA, SUITE 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1160 SO. STATE ST.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OREM, UT 84057</td>
</tr>
<tr>
<td>1</td>
<td>Dr. John B. Carroll</td>
<td>Psychometric Lab</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Univ. of N. Carolina</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Davie Hall 013A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chapel Hill, NC 27514</td>
</tr>
<tr>
<td>1</td>
<td>Charles Myers Library</td>
<td>Livingstone House</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Livingstone Road</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stratford</td>
</tr>
<tr>
<td></td>
<td></td>
<td>London E15 2LJ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ENGLAND</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Kenneth E. Clark</td>
<td>College of Arts &amp; Sciences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Rochester</td>
</tr>
<tr>
<td></td>
<td></td>
<td>River Campus Station</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rochester, NY 14627</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Norman Cliff</td>
<td>Dept. of Psychology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Univ. of So. California</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University Park</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Los Angeles, CA 90007</td>
</tr>
<tr>
<td>1</td>
<td>Dr. William E. Coffman</td>
<td>Director, Iowa Testing Programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>334 Lindquist Center</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Iowa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iowa City, IA 52242</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Meredith P. Crawford</td>
<td>American Psychological Association</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1200 17th Street, N.W.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Washington, DC 20036</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Fritz Drasgow</td>
<td>Yale School of Organization and Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yale University</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Box 1A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New Haven, CT 06520</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Mavin D. Dunnette</td>
<td>Personnel Decisions Research Institute</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2415 Foshay Tower</td>
</tr>
<tr>
<td></td>
<td></td>
<td>821 Marguette Avenue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minneapolis, MN 55402</td>
</tr>
<tr>
<td>1</td>
<td>Mike Durmeyer</td>
<td>Instructional Program Development</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Building 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NET-PDCD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Great Lakes NTC, IL 60088</td>
</tr>
<tr>
<td>1</td>
<td>ERIC Facility-Acquisitions</td>
<td>4833 Rugby Avenue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bethesda, MD 20014</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Benjamin A. Fairbank, Jr.</td>
<td>McFann-Gray &amp; Associates, Inc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5825 Callaghan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suite 225</td>
</tr>
<tr>
<td></td>
<td></td>
<td>San Antonio, Texas 78228</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Leonard Feldt</td>
<td>Lindquist Center for Measurement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Iowa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iowa City, IA 52242</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Richard L. Ferguson</td>
<td>The American College Testing Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.O. Box 168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iowa City, IA 52240</td>
</tr>
<tr>
<td>1</td>
<td>Dr. Victor Fields</td>
<td>Dept. of Psychology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Montgomery College</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rockville, MD 20850</td>
</tr>
<tr>
<td>1</td>
<td>Univ. Prof. Dr. Gerhard Fischer</td>
<td>Liebiggasse 5/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A 1010 Vienna</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AUSTRIA</td>
</tr>
</tbody>
</table>

23
1. Professor Donald Fitzgerald  
   University of New England  
   Armidale, New South Wales 2351  
   AUSTRALIA

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

3. Dr. John R. Frederiksen  
   Bolt Beranek & Newman  
   50 Houlton Street  
   Cambridge, MA 02138

4. Dr. Robert Glaser  
   LRDC  
   UNIVERSITY OF PITTSBURGH  
   3939 O'HARA STREET  
   PITTSBURGH, PA 15213

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

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

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

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

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

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

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

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

13. Dr. John A. Keats  
    University of Newcastle  
    AUSTRALIA 2308

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

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

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

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

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

19. Dr. Gary Marco  
    Educational Testing Service  
    Princeton, NJ 08450
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 Sympon  
Psychometric Research Group  
Educational Testing Service  
Princeton, NJ 08541

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

1 Dr. 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 Varie1 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