LEVEL

# Operational Characteristics of a One-Parameter
# Tailored Testing Procedure

Wayne M. Patience
and
Mark D. Reckase

Research Report 79-2
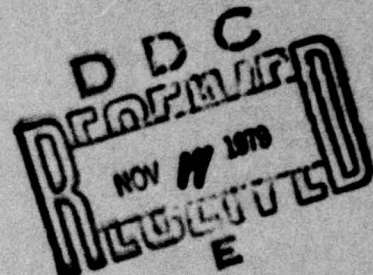October 1979

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

Um

79 14 11 044

(9) Research rept.,

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER RR-79-2 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) Operational Characteristics of a One-Parameter Tailored Testing Procedure, | | 5. TYPE OF REPORT & PERIOD COVERED Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) Wayne M. Patience and 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, Missouri 65201 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N  Proj.: RR042 T.A.: 042-04-01         04 W.V.: NR150-395 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217 | | 12. REPORT DATE October 1979 |
| | | 13. NUMBER OF PAGES 33 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

(16) RR04204

(17) RR0420401

16. DISTRIBUTION STATEMENT (of this Report)

Approval 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 | Rasch Model |
| Ability Testing | Tailored Testing |
| Latent Trait Models | Computerized Testing |

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

While numerous articles have appeared in the literature which describe the one-parameter logistic model and its application in a tailored testing setting, little or no research has been conducted on the operational characteristics of the procedure when program parameters and item pool attributes are varied. The primary objective of this investigation was to determine the effects of varying the program parameters, stepsize and acceptance range, as well as the item pool attributes, size and shape, on

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE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

410502

#20 (Cont.)

the bias and standard error of the maximum likelihood ability estimates obtained from tailored tests. Specifically, two main research questions were addressed. First, what values of stepsize and acceptance range provided the least bias and smallest standard error of ability estimates? The stepsize program parameter controlled the magnitude of movement through the item pool during the initial item selection phase of tailored testing. The acceptance range program parameter specified how deviant the selected item's difficulty value could be from the requested item difficulty and still be chosen for administration. Secondly, what shape and size of item difficulty distribution provided the least bias and standard error of ability estimates across the range of the latent trait? Two FORTRAN programs were used for investigating the effects of program parameters and item pool attributes. Both programs took as input the stepsize, acceptance range, item difficulty values for the various sizes and shapes of item pools, and the true abilities for which estimates were to be made. The first program, TREEIP, produced the propensity distribution, the probability distribution for observed ability estimates given a true ability, $\theta$, and provided output of the $E(\theta)$ and $\sqrt{VAR(\theta)}$. The other program, SIMIP, was developed to overcome the limitation on the size of item pool which could be investigated at a reasonable cost using the TREEIP program. The SIMIP program provided output of the $\bar{X}_{(\theta)}$ and $S.D._{(\theta)}$ of ability estimates of a specified number of simulated tailored tests assuming a given $\theta$. The results of the study were drawn from tables which summarized the output of the TREEIP and SIMIP programs. In addition to the recommendations regarding the research questions stated above, an effort was made to discuss the interaction of the variables of stepsize, acceptance range, item pool size and the shape of the distribution of item pool difficulties. Results suggested that each of these variables played a substantial role in affecting the magnitude of statistical bias and standard error at various points along the ability continuum. The results were presented as a guide for those involved in setting up a tailored testing procedure. The intent was to provide figures and tables to facilitate applications of tailored testing procedures such that a minimum of bias and standard error of ability estimates could be attained.

## CONTENTS

# Operational Characteristics of a One-Parameter

## Tailored Testing Procedure

Tailored testing, the selection and scoring of test items administered in an interactive fashion to individual examinees, has within the past decade become the spearhead for application of latent trait models to achievement and ability measurement. The availability of improved computer technology has contributed greatly to the increase in the number of systems presently in operation which administer tailored or adaptive tests. It should be noted that tailored testing as presented here is synonomous with many other assigned names such as adaptive testing, response contingent testing, or sequential testing. Of the many procedures available for tailored testing, one of those used at the University of Missouri is based on the one-parameter logistic model.

While numerous articles have appeared in the literature which describe the one-parameter logistic model and its application in a tailored testing setting (see, for example, Reckase, 1974; Weiss, 1974; Patience, 1977), little or no literature has been written discussing operational characteristics of the tailored testing procedure when program parameters and item pool attributes are varied. For this report, operational characteristics refer to how well the tailored testing procedure estimates a given true ability. Program parameters refer to those program options (such as the item selection rule) that must be selected before the program can operate. Item pool attributes refer to the size, distribution, and quality of the item pool. The operational characteristics, item pool attributes, and program parameters will be described in detail shortly.

Although no literature was found which addressed the effects of varying program parameters, a few studies have appeared in the literature which investigated effects of item pool attributes on the operation of tailored testing. Jensema (1975), for example, has investigated the influence of item pool size and item characteristics on a Bayesian tailored testing procedure. In general, Jensema found that when items are of adequate quality, it is not necessary to have very large item pools. Reckase (1976) concurred with Jensema in recommending a rectangular distribution of item pool difficulty values. In this latter study, the tailored testing procedure was based on an empirical maximum likelihood estimation of the ability parameter of the simple logistic (Rasch) model. Issues worthy of further investigation have surfaced in addition to item pool attributes, such as the effects of program parameters on the bias and variance of ability estimation.

Several articles have appeared in the literature which use the phrase "bias of tailored testing ability estimation" to mean procedural bias toward subgroups of an examinee population such as minorities (see, for

example, Pine and Weiss, 1978). The research reported here did not address this type of bias. Rather, ability estimate bias, as investigated by this paper, was concerned with whether the expected values of the maximum likelihood ability estimates were equal to the known true ability. In this sense, the attempt was to identify values for the program parameters and the item pool characteristics which would provide the least statistical bias in ability estimation. The variance of ability estimates was the squared standard error of the ability estimates for a known true ability. The desire was to minimize this standard error. These two dependent measures provided the criteria for judging how well the tailored testing procedure estimated known abilities when the program parameters and item pool characteristics were varied.

## Purpose

The primary purpose of the research described herein was to determine the operational characteristics of a one-parameter tailored testing procedure when program parameters and item pool attributes were varied. The program parameters investigated were the stepsize and acceptance range. The stepsize parameter specified the magnitude of movement of the ability estimate during the initial item selection phase of tailored testing. After the initial phase, maximum likelihood ability estimation was used. The acceptance range parameter determined how deviant the selected item's difficulty value could be from the requested item difficulty and still be acceptable for administration. In the tailored test, items were requested by the procedure to match the ability estimate computed based on previous item responses. The item pool attributes varied were size, shape, and quality. Each of these variables will now be described more specifically.

The premise of tailored testing is that when an examinee answers an item correctly, the next item administered should be more difficult, and when an examinee answers an item incorrectly, the next item should be less difficult. The stepsize program parameter initially controlled how much more difficult or easy was the next item administered. The selection of items was controlled by the fixed stepsize until the examinee had answered items both correctly and incorrectly. After both a correct and incorrect response had been obtained in the response string, a maximum likelihood ability estimate was obtained using an iterative search for the mode of the likelihood distribution. For a more complete description of the item selection and ability estimation components of this maximum likelihood tailored testing procedure see Patience (1977). In the past, arbitrary values have generally been chosen for the stepsize. One of the primary goals of this research was to empirically investigate the effects of stepsize values on the bias and standard error of ability estimates. In so doing, the intent was to determine the optimal stepsize value which would minimize the bias and standard error of ability estimates.

The second program parameter investigated was the acceptance range. The acceptance range specified the amount of deviation in difficulty an administered item could have from the requested item difficulty and still be acceptable for administration. The acceptance range parameter monitored the appropriateness of items selected throughout the tailored test, i.e.,

both during item selection based on the fixed stepsize until both correct and incorrect responses had been obtained, and also during item selection to maximize the information function for a maximum likelihood ability estimate. If more than one item were within plus or minus the acceptance range of the desired item, the item with a difficulty value nearest the requested value was chosen. If no item were available from the pool within the specified acceptance range of the difficulty requested, the tailored test was terminated. The primary aim regarding the acceptance range, then, was to determine what value or range of values yielded the least bias and standard error of ability estimates. Clearly, a small value for the acceptance range would have insured that items very near the desired item difficulty would be administered. On the other hand, too small an acceptance range value would have increased the chance of premature termination of the tailored test, which would have induced bias of the ability estimate. It should be noted that both stepsize and acceptance range interact with item pool attributes and, therefore, a choice of what values are optimal may not be made assuming independence of these controlling factors.

The item pool attributes studied in this research included size, shape, and quality. Simulated item pools used in this investigation ranged in size from nine to 181 items. Shapes of item pool distributions were normal, rectangular, bimodal, and skewed. Item pool quality referred to the contrast between actual and idealized pools. Idealized pools consisted of item difficulty parameters equally spaced from -3 to +3.

Actual pools consisted of item difficulty values (minus one times each of the item easiness values) obtained from calibration runs using the Wright and Panchapakesan (1969) calibration program based on the Rasch model. In these pools, items were not equally spaced on the difficulty scale. One of the actual item pools contained 72 items while the other had 180 items. The 72 item pool consisted of item difficulty parameter estimates from the calibration of three vocabulary tests. This pool was labeled VC1PL. The other pool was constructed using item difficulty parameter estimates from the calibrations of tests covering the evaluation techniques portion of an introductory measurement and evaluation course. This pool was labeled ET1PL. The distributions of item difficulty for VC1PL and ET1PL were graphed and appear in Appendix A. It should be noted that item pool attributes played a substantial role in the utility of the tailored testing procedure.

## Programs

Two FORTRAN programs were used for investigating effects of program parameters and item pool attributes. The input variables for both programs included: a) acceptance range, b) stepsize, c) item pool size, d) item difficulty values for the various sizes and shapes of item pools, and e) the true abilities for a set of hypothetical examinees. Both programs output the mean and standard deviation of the estimates of each true ability provided. These served as dependent measures for determination of the quality of estimation for the specific values of the acceptance range, stepsize, and item pool parameter set.

The first program, the TREEIP, was based on the concept of a propensity distribution. A propensity distribution in this context was defined as the probability distribution for observed ability estimates given a true ability, $P(\hat{\theta}|\theta)$ (Lord and Novick, 1968). The concept of a propensity distribution was extended from its use in true score theory to the context of latent trait ability estimation. The TREEIP program determined the propensity distribution for a given true ability, $\theta$, analytically from the properties of the tailored testing model.

Briefly, the TREEIP program operated as follows. Initially an item of average difficulty was administered to the simulated examinee with known true ability. Based on the probability function for the simple logistic model,

$$P(\mu) = \frac{e^{\mu(\theta - b)}}{1 + e^{(\theta - b)}} \tag{1}$$
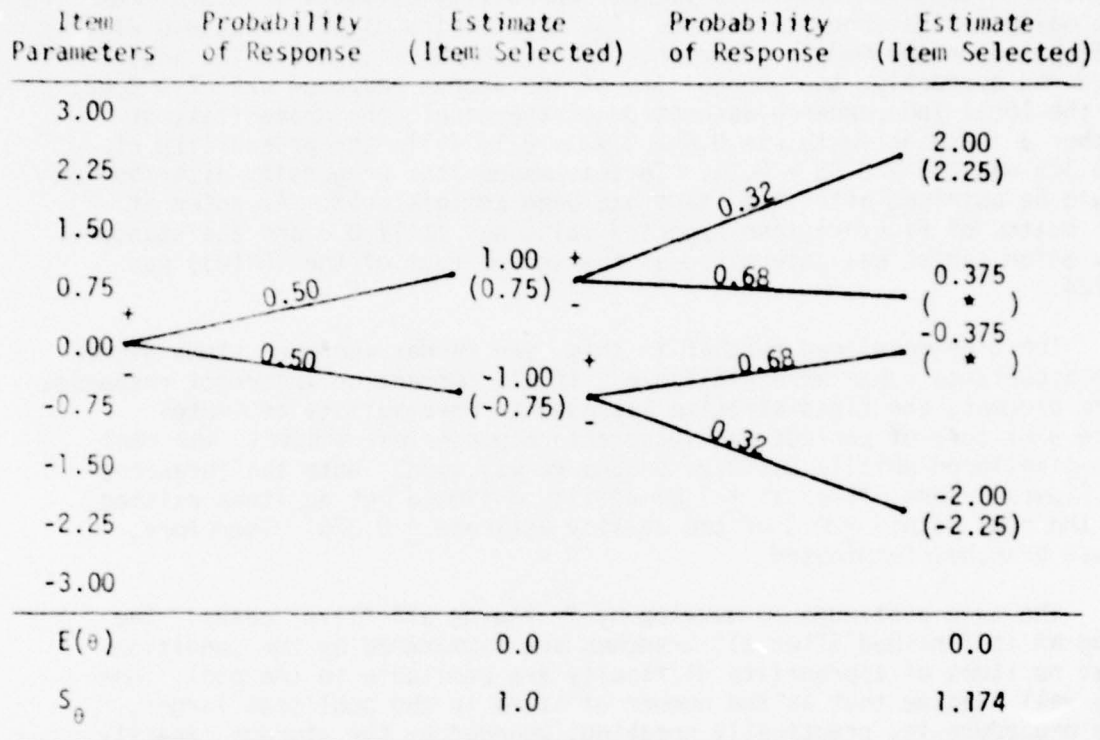
where $\mu$ is the item score (0 or 1), $b$ is the item difficulty parameter, and $\theta$ is the ability parameter, the probability of a correct and the probability of an incorrect response were obtained. If the response were correct, the ability estimate was increased by the stepsize. If the response were incorrect, the ability estimate was decreased by the stepsize. Thus after one item was administered, two paths or branches were present on the "tree". (The tree diagram from probability theory was employed to represent the propensity distribution in this study.) Based on these first possible ability estimates, the closest items to each of the two estimates was selected for administration with the constraint that the difficulty of the items must have been within plus or minus the acceptance range from the present ability estimates. If no items were available, that branch was terminated at that point. However, assuming items were available, there existed four possible paths after the second item had been administered. As long as all correct or all incorrect responses were obtained on a given path, the ability estimates continued to be increased or decreased, respectively, by the stepsize. However, when both a correct and an incorrect response were present on a particular path of the tree, a maximum-likelihood ability estimation procedure obtained an ability estimate using an iterative search for the mode of the likelihood distribution.

To partially illustrate how the propensity distribution was determined by the TREEIP, Figure 1 shows a diagram representing the operation of the procedure on a nine item rectangular pool. The stepsize used for this illustration was 1.0 and the acceptance range was 0.3. The $\theta$ for this analytical derivation of the propensity distribution was set at zero. As was pointed out above, the procedure began by administering an item of average difficulty from the pool, i.e., the item with the difficulty parameter 0.0. The probability of a correct response, as determined by the probability function given above for the simple logistic model, was 0.5 and the probability of an incorrect response was 0.5.

After a correct response the ability estimate was increased by the stepsize, or after an incorrect response, it was decreased by the stepsize. Thus after one item, the ability estimate was either 1.0 with

### Figure 1

Procedural Operation of TREE1P
on a Nine Item Pool with
Stepsize = 1.0 and Acceptance Range = 0.3

| Item Parameters | Probability of Response | Estimate (Item Selected) | Probability of Response | Estimate (Item Selected) |
|---|---|---|---|---|
| 3.00 | | | | |
| 2.25 | | | | 2.00 (2.25) |
| 1.50 | | | 0.32 | |
| 0.75 | 0.50 | 1.00 (0.75) + | 0.68 | 0.375 ( * ) |
| 0.00 + − | 0.50 | − | 0.68 | −0.375 ( * ) |
| −0.75 | | −1.00 (−0.75) + − | 0.32 | |
| −1.50 | | | | |
| −2.25 | | | | −2.00 (−2.25) |
| −3.00 | | | | |
| E(θ) | | 0.0 | | 0.0 |
| S_θ | | 1.0 | | 1.174 |

Note. The * indicates that no item was available in the pool within ± the acceptance range.

probability of 0.5 or −1.0 with a probability of 0.5. This procedure was followed so that finite ability estimates would be available after each item response, rather than the ± ∞ value given by the maximum-likelihood procedure. The expected value of the distribution after one item was 0.0 and the standard deviation was 1.0.

Based on these first possible ability estimates the closest items were selected from the pool with the restriction that their difficulties must have been within plus or minus 0.3 of the requested difficulties. Thus, as Figure 1 illustrates, items with parameter estimates of plus and minus 0.75 were administered to the estimated abilities plus and minus 1.00 respectively. On the upper branch of the tree, a correct response yielded an ability estimate that was again increased by the stepsize, since a maximum-likelihood estimate could not be determined without both a correct and incorrect response. Now, the ability estimate was 2.0. The probability of this correct response to the item with the 0.75 difficulty parameter was 0.32. The bottom branch of the tree was the same except for the change in sign of the item parameters and ability

estimates. When the item pool distribution being considered was symmetric, the results of the analyses were the same above the zero point as below the zero point except for the change in sign.

Following the middle branches of the tree, an incorrect response to the item with difficulty 0.75 yielded an ability estimate of 0.375 from the maximum-likelihood technique. The probability of this response was 0.68 based on the model. When the first item was missed and the second answered correctly, the probability of the second response was also 0.68. By the local independence assumption of the model, the probability of either a $\pm$ 2.0 estimate was 0.5 X 0.32 = 0.16 while the probability of $\pm$ 0.375 was 0.5 X 0.68 = 0.34. In this manner the propensity distribution could be obtained after two items had been administered. As noted at the bottom of Figure 1, the expected value was still 0.0 and the standard deviation (which was determined as the square root of the VAR($\theta$)) was 1.174.

The tree developed further in this same manner whenever items within the acceptance range were available. If all correct or incorrect responses were present, the fixed stepsize was used to make ability estimates. Once a mixture of correct and incorrect responses was present, the maximum-likelihood ability estimate procedure was used. Note the "branches" of Figure 1 were "live" at $\pm$ 2.00 ability estimate but no items existed in the pool within $\pm$ 0.3 of the ability estimate $\pm$ 0.375. Therefore, those branches terminated.

The tree continues to develop by following all "live" paths. The program is finished after all branches are terminated by the condition that no items of appropriate difficulty are available in the pool. One may well imagine that as the number of items in the pool gets larger, the procedure is, practically speaking, bounded by the storage capacity of the computer facility and magnitude of one's computer budget. For the IBM 370/168 system on which the TREE1P program was run, it was found that sixty-one items was the practical upper limit on the number of items the pool could contain for any particular run of the various combinations of stepsize, acceptance range, and shape of the item difficulty distribution.

Due to the limitation on size of the item pool which could be investigated with the TREE1P program, the second computer program, SIM1P, was developed. This program was adapted from the tailored testing procedure based on the Rasch model which was already operational. This particular tailored testing procedure has been described thoroughly elsewhere (Reckase, 1974), so only the details pertinent to this research have been presented. The SIM1P program followed only one path for any given $\theta$ in contrast to the TREE1P. A particular path was selected using Monte Carlo simulation techniques. It provided for investigation of the properties of bias and variance of ability estimation with much larger item pools since the required storage and computation were substantially reduced as compared to the TREE1P program.

The following values served as input to the program: the stepsize, acceptance range, item pool difficulty values, $\theta$, and number of simulated

tests to be administered by the tailored testing procedure. The procedure initially administered an item of average difficulty from the pool of items provided. If a correct response were obtained, the ability was increased by the stepsize. If an incorrect response were obtained, the ability was decreased by the stepsize. The appropriate item for the new ability was administered. This fixed stepsize up and down procedure continued until both a correct and incorrect answer had been obtained in the response string. Then the procedure switched from the fixed stepsize procedure to maximum-likelihood ability estimation. In both cases, items were selected to maximize the item information (Birnbaum, 1968). Ability estimation was accomplished after each item was administered (provided correct and incorrect responses had previously occurred) by the maximum-likelihood estimation procedure using an iterative search for the mode of the likelihood distribution. The items administered had to be within plus or minus the acceptance range from the requested item difficulty. If no items were available within this range of the estimated ability, the procedure stopped. The only other stopping rule was based on a preset maximum number of items that was to be administered.

Items were scored correct or incorrect by the SIMIP program utilizing an internal random number generator. First, the probability of a correct response was computed using the formula for the probability function of the simple logistic model stated earlier. The $\theta$ for this computation was the true $\theta$ that was input into the program, and the difficulty parameter, b, was that of the item just administered to the simulated examinee. After this probability of a correct response had been determined, the random number generator selected a number between zero and one from a rectangular distribution. If this randomly selected number was less than or equal to the probability of a correct response, the item was scored correct. If the randomly selected number was greater than the probability of a correct response, the item was scored as incorrect. An ability estimate was then obtained and the next item to be administered was selected to maximize information for this estimated ability. This procedure continued until one of the stopping rules was encountered.

The major controlling program parameters for both the TREEIP and SIMIP were the stepsize and acceptance range values. The stepsize parameter controlled how quickly the procedure would move through the item pool while the acceptance range parameter specified how discrepant items could be from those desired and still be administered. The acceptance range also indirectly determined the number of items from the pool which were available for administration. Clearly, the wider was the acceptance range, the greater was the number of items that could have been chosen for administration.

The TREEIP and SIMIP programs used in this study for determining the optimal stepsize, acceptance range, item pool size, and item pool distribution were similar in that both output the mean and standard deviation of ability estimated for each true $\theta$ input. However, they differed in the manner in which the mean and standard deviation were determined. While the TREEIP pursued all possible paths through the item pool, the SIMIP followed only the path that was the result of the simulated interaction of an examinee with the tailored testing procedure. The mean and

standard deviation from the TREEIP were actually expected values and square roots of variance computed from probabilities arising from the one-parameter model and ability estimates arising from the maximum-likelihood estimation technique. The SIMIP program provided a mean and standard deviation of the set of ability estimates obtained for each of the $\theta$'s specified.

## Research Design

To investigate the optimal stepsize, acceptance range, item pool size, and item pool shape, nearly all possible combinations of the following were input into the TREEIP and SIMIP programs for true abilities -3, -2, -1, 0, 1, 2, and 3. The stepsize values used were .3, .4, .5, .6, .693, .8, .9, 1.0, 1.5, 2.0, and 3.0, while acceptance ranges were .1, .2, .3, .4, and .5. Item pool sizes were 9, 13, 25, 31, 61, 72, 180, and 181. Item pool shapes investigated were normal, rectangular, bimodal, and skewed, with difficulty values constrained between plus and minus three. Idealized item pools (difficulty values in the above shapes with spacing dependent on shape and size of item pool) were constructed and used as input to the programs, as well as actual item pools (test items calibrated and formed into pools with no constraint on the spacing along the difficulty scale).

The manner in which item pool size effects were investigated using simulations was to run the TREEIP and SIMIP programs on the various sized pools mentioned above. With the resulting data, plots and projections were made to estimate the item pool sizes needed for various accuracies of ability estimation. The relationships between the item pool size, bias, and the standard deviation were determined.

The comparisons to determine the optimal combination of independent variables were based upon the mean and standard deviation of twenty-five simulated administrations of a tailored test to each $\theta$ using the SIMIP; where for the TREEIP program, the comparisons were of the expected value of $\theta$, $E(\theta)$, and the standard deviation of $\theta$, $\sqrt{Var(\theta)}$. Values of these dependent variables were compared across program runs using various sized item pools, holding stepsize and acceptance range constant. They were also compared from runs using various shapes of item pools, holding size of item pool, stepsize, and acceptance range fixed. Additionally, comparisons were made of the dependent variables, first varying stepsize with all other variables fixed, and then varying the value of the acceptance range while holding all other variables constant. Since the TREEIP program was considered to yield the most accurate values, i.e. $E(\theta)$ and $\sqrt{Var(\theta)}$ based upon the propensity distribution, another comparison was deemed important. Because the SIMIP means and standard deviations were subject to sample variation, they were validated against values of the TREEIP for various runs on the sixty-one item pool. Also, the number of estimates of the true ability, i.e. the number of tailored tests administered to each simulated examinee by the SIMIP program, was varied. This was done to check whether an appropriate number of administrations had been used.

## Results

The results of this study were to a great extent drawn from tables which summarized the results of the TREEIP and SIMIP programs. One issue to be investigated was the type of distribution of item pool difficulty parameters that yielded the least bias and standard error of ability estimates across the range of ability from -3 to +3. Another important question was how large an item pool was necessary to accomplish the goal of accurate ability estimation. Thirdly, a determination of the preferred magnitude of the stepsize parameter was desired. The fourth outcome of this study was to decide upon the approximate value of the acceptance range program parameter which would provide ability estimates with the least bias and standard error. These were the primary targets of the study.

Secondary goals of the study included a comparison of the performance of actual versus ideal item pools. Another secondary objective was to compare the results of the TREEIP and SIMIP programs. In this regard, two concerns were investigated. One pertained to how close the SIMIP estimates of the means and standard deviations of ability were to the $E(\theta)$ and $\sqrt{Var(\theta)}$ determined by the TREEIP. The importance of this particular concern related to how well the SIMIP analyses on larger item pools provided accurate data on the primary questions of this study. It should be recalled that the motivation for development of the SIMIP program was to investigate the research questions of the study on larger item pools than the TREEIP program would realistically accommodate. The second concern subsumed under comparison of the TREEIP and SIMIP programs was to decide whether or not 25 estimates of each ability by the SIMIP was an adequate number. Several analyses were run using the SIMIP program on various item pools from which data had already been obtained from the TREEIP. By running the SIMIP on these pools and holding all other variables fixed except the number of test administrations, data were obtained pertaining to the adequacy of the SIMIP estimates of the means and standard deviations. Another matter along this same line was investigated with runs of the SIMIP on some of the larger pools. This was the question of whether or not 20 items was an adequate upper limit on the number of items administered by the tailored test.

### Item Pool Shape

The TREEIP program (propensity distribution technique) was used to evaluate the effects of varying the shape of the item pool difficulty distribution on ability estimation. Four shapes of item pools were studied: rectangular, normal, bimodal and skewed. The rectangular item pools were obtained simply by selecting equally spaced items between +3.0 and -3.0 inclusive. The normal item pools were constructed such that the items were equally spaced in probability. That is, the area between item positions was kept constant in the range from +3.0 to -3.0 standard deviation units in the normal distribution. This procedure for producing the normally distributed pools had the effect of selecting more items around the difficulty value of zero and fewer items at the extremes. A similar procedure was used in selecting the item parameters for the bimodal pools as was

used for selecting the normal pools. The negative half of the pool was centered around -.693 and the area under the normal distribution was used to place items around this point up to zero and down to -3.0. The same was true for the positive half of the pool. The reason +.693 were chosen as the two modes of the bimodal distribution was that, prior to the construction of a bimodal pool, .693 had appeared promising as a stepsize value. Therefore, after the first item was administered at 0, the stepsize of .693 would move the ability estimate out to one of the more dense regions of the pool depending upon whether the examinee correctly or incorrectly answered the first item. The skewed item pool distribution of item parameters was constructed via a similar procedure to that for the normal and bimodal pools. That is, the items divided the distribution into equal areas. For the skewed pool, tables of the Pearson Type III distribution were used. The pool constructed was positively skewed (skewness = .5). It should be noted that in the tables included in this report, a skewed distribution always indicates a positive skew. However, the results would generalize to negatively skewed pools.

Results concerning the shape of the item pool distribution may be seen in Tables 1-6 for different combinations of values of the other variables. However, Tables 1 and 2 point out the more general trends of the item distribution study. In Table 1 the comparisons of the normal and rectangular pools of 25 items are shown for only acceptance ranges of 0.1 and 0.3 when paired with stepsizes of 0.5 and 0.7 respectively. These values of acceptance range and stepsize were chosen because they appeared to yield some of the least bias and least variance estimates. Specifically, the acceptance range of 0.1 was chosen to check whether the more dense item parameters near the middle of the normal distribution would make the use of the smaller acceptance range desirable.

Table 1
Comparison of TREE1P Results from
25 Item Rectangular and Normal Item Distributions

| Acceptance Range | Step Size | Distribution Shape | Ability Level | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.0 | | 0.5 | | 1.0 | | 2.0 | | 3.0 | |
| | | | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| 0.1 | 0.5 | R | -0.001 | 0.918 | 0.470 | 0.927 | 0.944 | 0.943 | 1.893 | 0.968 | 2.764 | 0.884 |
| | | N | -0.009 | 0.951 | 0.522 | 0.904 | 0.980 | 0.762 | 1.468 | 0.426 | 1.555 | 0.251 |
| 0.3 | 0.7 | R | -0.013 | 0.787 | 0.430 | 0.824 | 0.911 | 0.893 | 1.986 | 0.984 | 2.933 | 0.773 |
| | | N | -0.000 | 0.959 | 0.623 | 0.922 | 1.169 | 0.821 | 1.877 | 0.491 | 2.093 | 0.231 |

As can be seen from Table 1, the normal distribution appears to be inferior to the rectangular item distribution in almost all cases. Except for the 0.1 acceptance range data at 0.5 and 1.0 ability levels, either

the expected values deviate more from the true θ or the standard deviations are larger, or both. It is interesting to note that even the estimates at ability 0.0 are not as good for the normally distributed pool as for the rectangular pool, even though more items are present for estimation of ability at that point in the normal pool.

Table 2
Expected Values and Standard Deviations
from TREEIP on Various Shaped Item Pools

| Pool Shape | Ability Level | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | | 1 | | 2 | | 3 | |
| | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| normal | -0.000 | 0.866 | 1.256 | 0.873 | 2.267 | 0.693 | 2.840 | 0.543 |
| bimodal | -0.000 | 0.857 | 1.245 | 0.876 | 2.281 | 0.688 | 2.852 | 0.525 |
| rectangular | -0.001 | 0.610 | 1.008 | 0.677 | 2.138 | 0.805 | 3.111 | 0.566 |
| skewed | 0.040 | 0.815 | 1.282 | 0.858 | 2.257 | 0.670 | 2.801 | 0.561 |

Note. All runs were on pools with 61 items with the stepsize and acceptance range program parameters set at 0.693 and 0.30 respectively. Only positive abilities were presented since the results are symmetric around zero except for the skewed pool.

Table 2 shows the expected values and standard deviations from the TREEIP on normal, bimodal, rectangular, and positively skewed pools. Each of these contained sixty-one items. The stepsize was fixed at 0.693, and the acceptance range was held at 0.30 for all runs. Again the rectangular pool performed better overall than did the other shapes of item difficulty distributions. For true abilities zero and one, the standard deviation of ability estimates, as well as the bias of the estimates, was generally small for analyses using the rectangular pool. At the ability levels of two and three, the rectangular pool yielded estimates with less bias in the expected values but larger standard deviations than the other shaped pools.

The results obtained from the TREEIP would have been the same for the negative end of the ability continuum when the pools were symmetric. Therefore, only the positive values of ability were run for the normal, rectangular, and bimodal pools. However, for the skewed pool containing sixty-one items, the negative ability values of -1, -2, and -3 were run using the same program parameters as were indicated in Table 2. The results were as follows. For -1, the $E(\theta)$ = -1.189 and $S_\theta$ = 0.836. For -2, the $E(\theta)$ = -2.249 and $S_\theta$ = 0.761. For -3, the $E(\theta)$ = -2.935 and $S_\theta$ = 0.577. Even if one considered this skewed pool as being better suited for ability levels around minus two to minus one, since it contained more items around that region, it did not perform better than the rectangular pool.

## Item Pool Size

The criteria for judging how large an item pool was needed for good ability estimation using the tailored testing procedure were again the bias and standard error of ability estimates. The results of the simulations using both the TREE1P and SIM1P programs have been condensed, and the general trend has been illustrated in Figure 2. The values of the $E(\theta)$ and $S_\theta$ which have been plotted for item pools of size 9, 13, 25, 31, and 61 were obtained from the TREE1P. Each of these pools had a rectangular distribution of item difficulty parameters. The means and standard deviations of ability estimates on the SIM1P runs on VC1PL and ET1PL (described earlier) have been included in the plots of Figure 2. Each analysis represented in this figure had $\theta$ set equal to 1.0, the stepsize fixed at 0.693, and acceptance range equal to 0.30.

The top graph of Figure 2 illustrates that as item pool size reaches 61 for this particular set of analyses, the $E(\theta)$ is equal to $\theta$. The bias of the ability estimates is essentially zero. The bottom graph of Figure 2 shows that as item pool size increases, the standard error decreases. While these plots should be considered as rough approximations of the relationship between item pool size and ability estimate bias and standard error, the indication appears to be that with a uniform distribution of item difficulty, $\theta = 1$, and the program parameters equal to the values used here, one could expect very little bias and a standard error of about 0.3 with an item pool consisting of around 200 items. More will be presented on item pool size in the discussion section of this report.

## Stepsize

The results of the study of the preferred magnitude of the stepsize program parameter may be seen in Tables 3, 4, 5, and 7. Tables 3, 4, and 5 give the $E(\theta)$ and $S_\theta$ from TREE1P analyses of $\theta$ = 0, 1, 2, and 3 using item pools of size 9, 13, 25, 31, and 61 for the rectangular, normal, and bimodal distributions of item difficulty parameters, respectively. Table 7 presents the results of the SIM1P analyses on the ET1PL item pool for $\theta$ = -3, -2, -1, 0, 1, 2, and 3. Negative $\theta$ values are not shown in Tables 3, 4, and 5 since the results of the TREE1P on the pools used are the same as for the positive $\theta$ values except for the change of sign. This was expected since the item pool distributions of item difficulty are symmetric around zero. The acceptance range for all analyses for Tables 3, 4, and 5 was 0.30. For the SIM1P analyses of the ET1PL, a substantially larger item pool, a smaller acceptance range, 0.25, was used as is noted at the bottom of Table 7. Another variable recorded in Table 7 is the mean number of items administered for the 25 tests simulated by the SIM1P for each ability level. The maximum number of items per simulated test was 20 for these SIM1P analyses.

In general, results presented in Tables 3, 4, and 5 suggest that stepsizes between 0.5 and 1.0 give fairly unbiased estimates, and also have the smallest standard errors. Larger stepsizes tend to have a positive bias and larger standard errors. From several graphs like the ones presented in Figure 3, the stepsize value of 0.693 appears to be the best

Figure 2
Relationship Between Item Pool Size
and the $E(\theta)$ and $S_\theta$
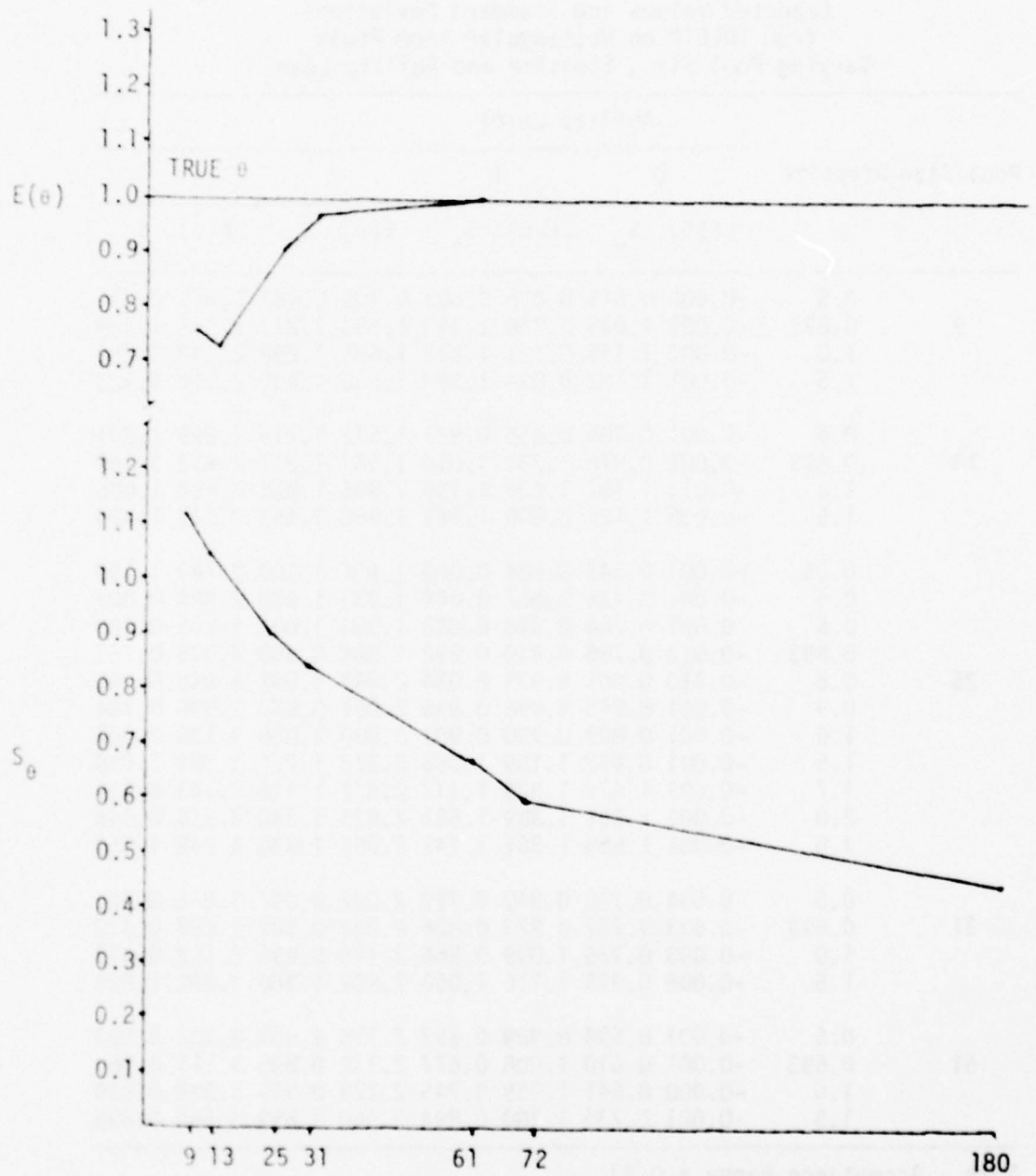


Item Pool Size

$\theta = 1$          Stepsize = 0.693          Acceptance Range = 0.30

Table 3
Expected Values and Standard Deviations
from TREE1P on Rectangular Item Pools
Varying Pool Size, Stepsize and Ability Level

| Pool Size | Stepsize | Ability Level | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | 0 | | 1 | | 2 | | 3 | |
| | | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| 9 | 0.5 | -0.000 | 0.645 | 0.405 | 0.603 | 0.709 | 0.482 | 0.877 | 0.335 |
| | 0.693 | -0.001 | 1.025 | 0.756 | 1.113 | 1.593 | 1.217 | 2.388 | 1.139 |
| | 1.0 | -0.001 | 1.155 | 0.821 | 1.213 | 1.685 | 1.298 | 2.548 | 1.286 |
| | 1.5 | -0.001 | 1.182 | 0.934 | 1.268 | 1.966 | 1.439 | 3.016 | 1.423 |
| 13 | 0.5 | -0.001 | 0.765 | 0.655 | 0.937 | 1.577 | 1.219 | 2.599 | 1.201 |
| | 0.693 | -0.001 | 0.976 | 0.733 | 1.056 | 1.587 | 1.217 | 2.454 | 1.168 |
| | 1.0 | -0.001 | 1.187 | 1.037 | 1.150 | 1.995 | 1.085 | 2.822 | 1.005 |
| | 1.5 | -0.006 | 1.125 | 0.899 | 1.249 | 1.960 | 1.463 | 3.045 | 1.424 |
| 25 | 0.25 | -0.001 | 0.547 | 0.584 | 0.809 | 1.606 | 1.200 | 2.783 | 1.190 |
| | 0.5 | -0.001 | 0.736 | 0.857 | 0.842 | 1.933 | 1.000 | 2.964 | 0.809 |
| | 0.6 | 0.001 | 0.744 | 0.896 | 0.888 | 1.986 | 1.004 | 2.955 | 0.788 |
| | 0.693 | -0.013 | 0.786 | 0.910 | 0.892 | 1.984 | 0.980 | 2.925 | 0.765 |
| | 0.8 | -0.013 | 0.801 | 0.931 | 0.934 | 2.047 | 1.042 | 3.045 | 0.845 |
| | 0.9 | -0.001 | 0.845 | 0.996 | 0.895 | 2.061 | 0.972 | 2.996 | 0.784 |
| | 1.0 | -0.001 | 0.829 | 0.990 | 0.901 | 2.099 | 1.036 | 3.135 | 0.867 |
| | 1.5 | -0.001 | 0.972 | 1.109 | 1.086 | 2.318 | 1.221 | 3.389 | 1.040 |
| | 1.7 | -0.001 | 1.473 | 1.329 | 1.417 | 2.477 | 1.116 | 3.143 | 0.614 |
| | 2.0 | -0.001 | 1.551 | 1.389 | 1.553 | 2.673 | 1.348 | 3.535 | 0.846 |
| | 3.0 | -0.001 | 1.555 | 1.361 | 1.741 | 2.863 | 1.930 | 4.248 | 1.750 |
| 31 | 0.5 | 0.004 | 0.726 | 0.949 | 0.788 | 2.022 | 0.902 | 3.018 | 0.725 |
| | 0.693 | -0.003 | 0.742 | 0.973 | 0.826 | 2.068 | 0.907 | 2.997 | 0.672 |
| | 1.0 | -0.003 | 0.776 | 1.009 | 0.866 | 2.140 | 0.995 | 3.183 | 0.817 |
| | 1.5 | -0.005 | 0.925 | 1.116 | 1.050 | 2.002 | 1.388 | 3.382 | 1.023 |
| 61 | 0.5 | -0.001 | 0.598 | 0.989 | 0.657 | 2.116 | 0.804 | 3.133 | 0.593 |
| | 0.693 | -0.001 | 0.610 | 1.008 | 0.677 | 2.138 | 0.805 | 3.111 | 0.566 |
| | 1.0 | -0.000 | 0.641 | 1.039 | 0.745 | 2.229 | 0.915 | 3.239 | 0.689 |
| | 1.5 | -0.001 | 0.734 | 1.100 | 0.894 | 3.560 | 0.899 | 3.560 | 0.899 |

Note. Acceptance Range = 0.30

Table 4
Expected Values and Standard Deviations
from TREE1P on Normal Item Pools
Varying Pool Size, Stepsize and Ability Level

| Pool Size | Stepsize | Ability Level | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0 | | 1 | | 2 | | 3 | |
| | | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| 9 | 0.5 | -0.001 | 1.018 | 0.848 | 0.847 | 1.318 | 0.491 | 1.463 | 0.226 |
| | 0.693 | -0.001 | 1.098 | 0.960 | 0.966 | 1.601 | 0.655 | 1.898 | 0.382 |
| | 1.0 | -0.001 | 1.269 | 0.880 | 1.084 | 1.641 | 0.632 | 1.877 | 0.334 |
| | 1.5 | 0.000 | 1.500 | 0.693 | 1.330 | 1.142 | 0.972 | 1.358 | 0.638 |
| 13 | 0.5 | -0.001 | 1.028 | 1.062 | 0.866 | 1.697 | 0.514 | 1.922 | 0.237 |
| | 0.693 | -0.001 | 1.101 | 1.002 | 0.942 | 1.648 | 0.628 | 1.932 | 0.358 |
| | 1.0 | -0.000 | 1.273 | 1.146 | 1.020 | 1.760 | 0.548 | 1.946 | 0.231 |
| | 1.5 | -0.001 | 1.439 | 1.272 | 1.282 | 2.219 | 1.188 | 3.031 | 1.258 |
| 25 | 0.25 | -0.001 | 0.847 | 1.110 | 0.858 | 1.969 | 0.576 | 2.210 | 0.408 |
| | 0.5 | -0.001 | 0.891 | 1.184 | 0.837 | 2.016 | 0.572 | 2.359 | 0.278 |
| | 0.6 | -0.001 | 0.980 | 1.203 | 0.847 | 1.965 | 0.528 | 2.263 | 0.266 |
| | 0.693 | -0.000 | 0.956 | 1.174 | 0.811 | 1.871 | 0.482 | 2.079 | 0.227 |
| | 0.8 | -0.001 | 1.009 | 1.234 | 0.871 | 2.004 | 0.539 | 2.292 | 0.253 |
| | 0.9 | -0.001 | 1.052 | 1.290 | 0.964 | 2.223 | 0.784 | 2.818 | 0.658 |
| | 1.0 | -0.001 | 1.055 | 1.295 | 0.979 | 2.263 | 0.858 | 2.949 | 0.820 |
| | 1.5 | -0.001 | 1.327 | 1.384 | 1.186 | 2.394 | 1.070 | 3.167 | 1.114 |
| | 1.7 | -0.001 | 1.536 | 1.521 | 1.363 | 2.549 | 0.968 | 3.047 | 0.628 |
| | 2.0 | -0.001 | 1.738 | 1.653 | 1.600 | 2.845 | 1.248 | 3.492 | 0.884 |
| | 3.0 | -0.001 | 1.792 | 1.627 | 1.749 | 2.928 | 1.814 | 4.045 | 1.883 |
| 31 | 0.5 | -0.000 | 0.869 | 1.218 | 0.805 | 2.046 | 0.557 | 2.385 | 0.277 |
| | 0.693 | -0.001 | 0.964 | 1.268 | 0.880 | 2.192 | 0.734 | 2.778 | 0.607 |
| | 1.0 | -0.001 | 1.018 | 1.323 | 0.951 | 2.300 | 0.823 | 2.969 | 0.787 |
| | 1.5 | -0.001 | 1.301 | 1.404 | 1.155 | 2.410 | 1.043 | 3.176 | 1.092 |
| 61 | 0.5 | -0.000 | 0.753 | 1.201 | 0.797 | 2.132 | 0.541 | 2.465 | 0.254 |
| | 0.693 | -0.000 | 0.866 | 1.256 | 0.873 | 2.267 | 0.693 | 2.840 | 0.543 |
| | 1.0 | -0.000 | 0.915 | 1.298 | 0.944 | 2.361 | 0.774 | 3.010 | 0.711 |
| | 1.5 | -0.000 | 1.232 | 1.399 | 1.141 | 2.473 | 1.004 | 3.227 | 1.044 |

Note. Acceptance Range = 0.30

Table 5
Expected Values and Standard Deviations
from TREE1P on Bimodal Item Pools
Varying Pool Size, Stepsize and Ability Level

| Pool Size | Stepsize | Ability Level 0 | | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|---|---|---|
| | | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| 9 | 0.5 | -0.004 | 1.020 | 0.231 | 0.443 | 1.312 | 0.495 | 1.473 | 0.245 |
| | 0.693 | -0.004 | 1.095 | 0.951 | 0.968 | 1.601 | 0.666 | 1.903 | 0.383 |
| | 1.0 | -0.001 | 1.264 | 1.036 | 1.042 | 1.639 | 0.628 | 1.876 | 0.331 |
| | 1.5 | -0.001 | 1.442 | 1.216 | 1.326 | 2.187 | 1.252 | 3.027 | 1.291 |
| 13 | 0.5 | -0.001 | 1.006 | 1.009 | 0.903 | 1.671 | 0.579 | 1.917 | 0.275 |
| | 0.693 | -0.001 | 1.104 | 1.001 | 0.945 | 1.647 | 0.630 | 1.932 | 0.358 |
| | 1.0 | -0.000 | 1.267 | 1.143 | 1.011 | 1.754 | 0.551 | 1.946 | 0.238 |
| | 1.5 | -0.000 | 1.436 | 1.274 | 1.276 | 2.217 | 1.181 | 3.029 | 1.252 |
| 25 | 0.25 | -0.000 | 0.920 | 1.102 | 0.855 | 2.001 | 0.623 | 2.264 | 0.421 |
| | 0.5 | -0.001 | 0.870 | 1.152 | 0.867 | 2.024 | 0.594 | 2.373 | 0.278 |
| | 0.6 | -0.001 | 0.951 | 1.173 | 0.875 | 1.976 | 0.536 | 2.271 | 0.242 |
| | 0.693 | -0.001 | 0.964 | 1.207 | 0.933 | 2.174 | 0.768 | 2.774 | 0.612 |
| | 0.8 | -0.001 | 0.953 | 1.183 | 0.887 | 2.020 | 0.589 | 2.335 | 0.272 |
| | 0.9 | -0.002 | 1.025 | 1.260 | 0.994 | 2.246 | 0.780 | 2.833 | 0.631 |
| | 1.0 | -0.001 | 1.017 | 1.257 | 1.002 | 2.280 | 0.860 | 2.969 | 0.791 |
| | 1.5 | -0.001 | 1.294 | 1.350 | 1.192 | 2.396 | 1.064 | 3.176 | 1.091 |
| | 1.7 | 0.002 | 1.491 | 1.483 | 1.362 | 2.543 | 0.959 | 3.047 | 0.612 |
| | 2.0 | -0.000 | 1.717 | 1.609 | 1.592 | 2.831 | 1.235 | 3.485 | 0.871 |
| | 3.0 | -0.001 | 1.761 | 1.601 | 1.763 | 2.953 | 1.803 | 4.070 | 1.857 |
| 31 | 0.5 | -0.000 | 0.796 | 1.145 | 0.816 | 2.060 | 0.621 | 2.476 | 0.406 |
| | 0.693 | -0.000 | 0.924 | 1.229 | 0.912 | 2.218 | 0.741 | 2.814 | 0.585 |
| | 1.0 | -0.000 | 0.957 | 1.262 | 0.956 | 2.298 | 0.832 | 3.004 | 0.758 |
| | 1.5 | -0.002 | 0.968 | 1.284 | 1.049 | 2.446 | 1.080 | 3.338 | 1.015 |
| 61 | 0.5 | 0.006 | 0.726 | 1.174 | 0.800 | 2.246 | 0.692 | 2.903 | 0.572 |
| | 0.693 | -0.000 | 0.857 | 1.245 | 0.876 | 2.281 | 0.688 | 2.852 | 0.525 |
| | 1.0 | 0.033 | 0.867 | 1.221 | 0.897 | 2.356 | 0.820 | 3.107 | 0.714 |
| | 1.5 | 0.185 | 1.128 | 1.249 | 1.003 | 2.497 | 1.050 | 3.407 | 0.949 |

Note. Acceptance Range = 0.30

Table 6
Means and Standard Deviations
from SIM1P on a Bimodal and
Skewed Item Pool Varying
Number of Test Administrations

| Number of Tests Administered | Shape of Pool | | | |
|---|---|---|---|---|
| | Bimodal | | Skewed | |
| | $\overline{X}_\theta$ | $S_\theta$ | $\overline{X}_\theta$ | $S_\theta$ |
| 25 | 2.207 | 0.627 | 2.193 | 0.622 |
| 50 | 2.242 | 0.634 | 2.225 | 0.627 |
| 75 | 2.262 | 0.645 | 2.216 | 0.603 |

Note. All runs made with 20 item upper limit, stepsize
= .693, and acceptance range = 0.30. The true
ability was set at 2.0. Both the pools had 61
items.

Table 7
Means and Standard Deviations
from SIM1P on ET1PL Item Pool
Varying Stepsize

| Stepsize | | Ability Level | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| .1 | $\overline{X}_\theta$ | -2.886 | -2.145 | -0.992 | -0.050 | 1.135 | 1.991 | 3.331 |
| | $S_\theta$ | 0.715 | 0.728 | 0.486 | 0.534 | 0.502 | 0.627 | 0.788 |
| | Mni* | 13.04 | 15.88 | 19.24 | 20.00 | 20.00 | 19.84 | 18.40 |
| .2 | $\overline{X}_\theta$ | -2.779 | -2.230 | -1.132 | 0.129 | 0.952 | 2.009 | 2.972 |
| | $S_\theta$ | 0.491 | 0.681 | 0.550 | 0.461 | 0.374 | 0.515 | 0.857 |
| | Mni* | 12.24 | 13.96 | 19.68 | 20.00 | 20.00 | 19.76 | 18.24 |
| .3 | $\overline{X}_\theta$ | -3.157 | -2.139 | -1.134 | 0.064 | 1.018 | 2.055 | 3.213 |
| | $S_\theta$ | 0.652 | 0.645 | 0.800 | 0.503 | 0.363 | 0.516 | 0.844 |
| | Mni* | 10.04 | 14.48 | 18.56 | 20.00 | 19.92 | 19.56 | 16.08 |
| .4 | $\overline{X}_\theta$ | -3.168 | -2.250 | -1.052 | 0.001 | 1.070 | 1.987 | 2.910 |
| | $S_\theta$ | 0.611 | 0.782 | 0.547 | 0.518 | 0.444 | 0.531 | 0.554 |
| | Mni* | 9.56 | 17.04 | 19.24 | 20.00 | 20.00 | 19.48 | 18.12 |
| .5 | $\overline{X}_\theta$ | -2.762 | -2.096 | -1.122 | -0.070 | 1.136 | 2.076 | 3.053 |
| | $S_\theta$ | 0.539 | 0.619 | 0.700 | 0.539 | 0.562 | 0.548 | 0.718 |
| | Mni* | 9.20 | 14.72 | 18.12 | 20.00 | 20.00 | 19.40 | 16.28 |

Note. All runs made with 25 administrations per ability level, 20 item upper
limit, and acceptance range = .25.
*Mni = mean number of items administered.

Table 7 (Cont.)
Means and Standard Deviations
from SIMIP on ET1PL Item Pool
Varying Stepsize

| Stepsize | | Ability Level | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| .7 | $\overline{X}_\theta$ | -3.061 | -2.175 | -1.026 | -0.065 | 1.029 | 1.950 | 2.913 |
| | $S_\theta$ | 0.561 | 0.460 | 0.573 | 0.469 | 0.516 | 0.696 | 0.533 |
| | Mni* | 7.80 | 13.12 | 18.92 | 20.00 | 20.00 | 19.16 | 15.92 |
| .9 | $\overline{X}_\theta$ | -3.134 | -2.271 | -1.241 | 0.094 | 0.959 | 2.029 | 3.310 |
| | $S_\theta$ | 0.499 | 0.790 | 0.898 | 0.419 | 0.380 | 0.531 | 0.799 |
| | Mni* | 5.92 | 11.40 | 16.96 | 20.00 | 19.84 | 19.20 | 13.28 |
| 1.5 | $\overline{X}_\theta$ | -3.739 | -2.501 | -1.389 | 0.101 | 1.035 | 2.437 | 3.239 |
| | $S_\theta$ | 0.876 | 0.961 | 0.910 | 0.598 | 0.792 | 1.118 | 1.010 |
| | Mni* | 5.32 | 10.80 | 18.04 | 20.00 | 19.32 | 16.16 | 12.72 |
| 2.0 | $\overline{X}_\theta$ | -3.683 | -2.972 | -1.482 | -0.329 | 1.100 | 2.032 | 3.631 |
| | $S_\theta$ | 0.514 | 1.044 | 1.194 | 1.175 | 0.450 | 0.913 | 1.345 |
| | Mni* | 4.24 | 8.76 | 16.56 | 18.56 | 19.96 | 18.48 | 13.36 |
| 3.0 | $\overline{X}_\theta$ | -4.530 | -2.942 | -1.751 | -0.042 | 1.230 | 2.511 | 4.471 |
| | $S_\theta$ | 1.591 | 1.494 | 1.916 | 0.465 | 1.117 | 1.556 | 1.519 |
| | Mni* | 5.04 | 10.68 | 16.52 | 20.00 | 19.28 | 17.04 | 8.60 |

<u>Note</u>. All runs made with 25 administrations per ability level, 20 item upper
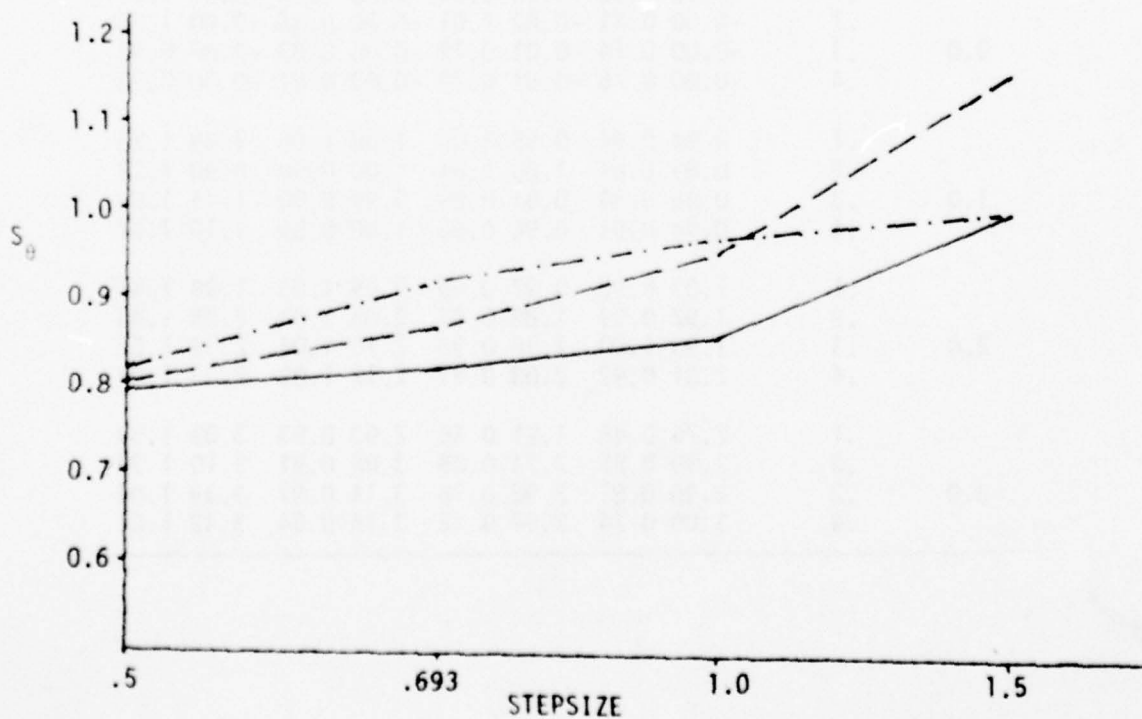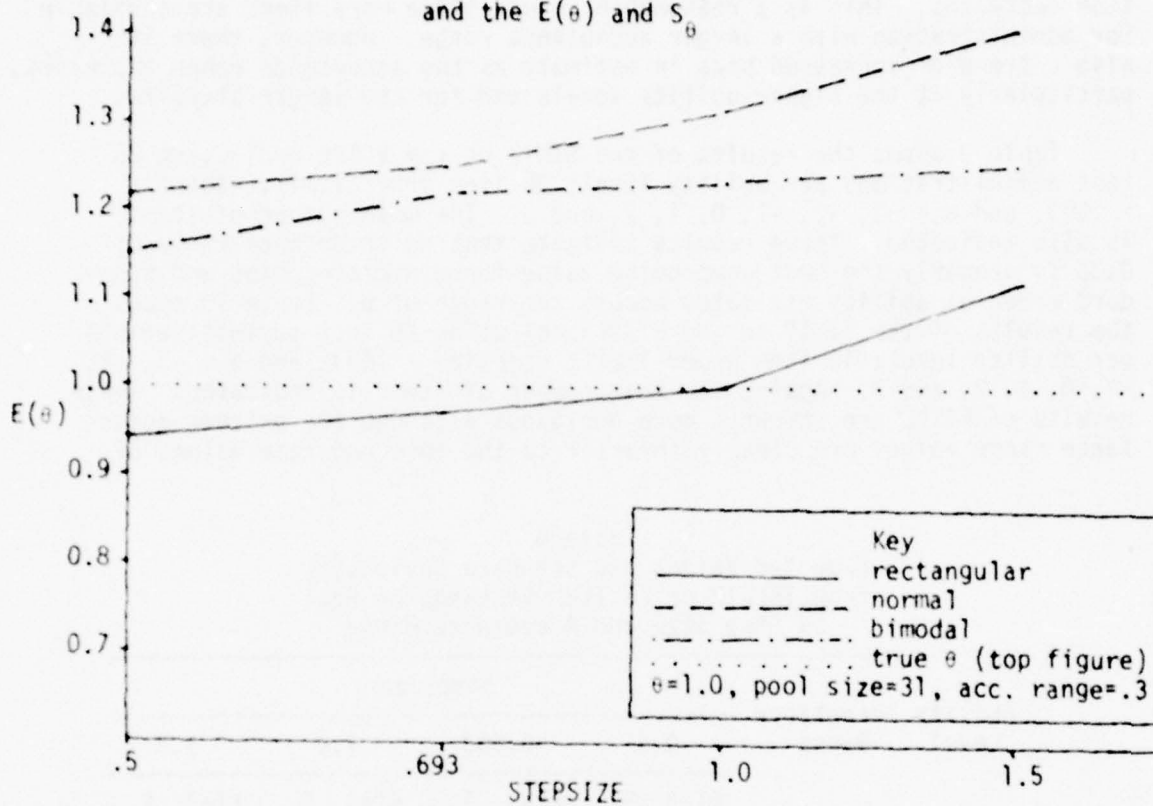limit, and acceptance range = .25.
*Mni = mean number of items administered

overall compromise value which achieves less bias while holding the standard error down. Figure 3 shows the $E(\theta)$ and $S_\theta$ for the 31 item rectangular, normal and bimodal pools when $\theta$ = 1.0 and the acceptance range equals 0.30 for various stepsizes.

Table 7, which reports the results of the SIMIP on the ET1PL pool, presents information that suggests a stepsize between 0.4 and 0.7 yields less bias and a smaller standard error. It should be recalled that the SIMIP is subject to sample variation, but in general, the results seem to suggest that a stepsize of about 0.7 is appropriate. However, a trend which should be investigated further is that larger item pools seem to do better with smaller stepsizes and conversely.

## Acceptance Range

The results of the acceptance range study are given in Tables 8, 9, and 10. Table 8 presents the $E(\theta)$ and $S_\theta$ for stepsizes 0.5, 0.693, 1.0, and 1.5; acceptance ranges 0.1, 0.2, 0.3, and 0.4; and ability levels 0.0, 1.0, 2.0, and 3.0 from TREEIP analyses. All of the results in Table

Figure 3
Relationship Between Stepsize
and the $E(\theta)$ and $S_\theta$

8 are based on the 25 item rectangular pool. From Table 8 it can be seen
that in most cases, as the acceptance range increases, the standard devia-
tion decreases. This is a reasonable result since more items are available
for administration with a larger acceptance range. However, there is
also a trend of increased bias in estimate as the acceptance range increases,
particularly at the higher ability levels and for the larger stepsizes.

Table 9 shows the results of the SIM1P on the VC1PL pool using 25
test administrations per ability level; 20 item upper limit; stepsize
= .693; and $\theta$ = -3, -2, -1, 0, 1, 2, and 3. The mean number of items
is also indicated. These results indicate that an acceptance range of
0.30 is probably the best compromise value for minimizing bias and stan-
dard error of ability estimates across the range of $\theta$. Table 10 shows
the results of the SIM1P on the ET1PL pool using 25 test administrations
per ability level; 40 item upper limit; stepsize = .693; and $\theta$ = -3, -2,
-1, 0, 1, 2, and 3. Again, the mean number of items is indicated. These
results on ET1PL are somewhat more ambiguous although the extreme accep-
tance range values are clearly inferior to the more moderate values of

Table 8
Expected Values and Standard Deviations
from TREE1P on 25 Item Rectangular Pool
by Step Size and Acceptance Range

| Ability Level | Acceptance Range | Stepsize | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0.5 | | 0.693 | | 1.0 | | 1.5 | |
| | | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ | $E(\theta)$ | $S_\theta$ |
| 0.0 | .1 | -0.00 | 0.92 | -0.00 | 0.84 | -0.00 | 1.04 | -0.01 | 1.07 |
| | .2 | -0.00 | 0.81 | -0.02 | 1.01 | -0.00 | 0.88 | -0.00 | 1.06 |
| | .3 | -0.00 | 0.74 | -0.01 | 0.79 | -0.00 | 0.83 | -0.00 | 0.97 |
| | .4 | -0.00 | 0.76 | -0.01 | 0.78 | -0.00 | 0.81 | -0.00 | 0.93 |
| 1.0 | .1 | 0.94 | 0.94 | 0.55 | 0.80 | 1.08 | 1.06 | 0.89 | 1.23 |
| | .2 | 0.89 | 0.87 | 1.00 | 1.04 | 1.00 | 0.94 | 0.90 | 1.22 |
| | .3 | 0.86 | 0.84 | 0.91 | 0.89 | 0.99 | 0.90 | 1.11 | 1.09 |
| | .4 | 0.94 | 0.81 | 0.96 | 0.83 | 1.00 | 0.89 | 1.10 | 1.07 |
| 2.0 | .1 | 1.89 | 0.97 | 0.97 | 0.66 | 2.09 | 1.03 | 1.99 | 1.45 |
| | .2 | 1.92 | 0.99 | 1.88 | 0.97 | 2.08 | 1.04 | 2.00 | 1.45 |
| | .3 | 1.93 | 1.00 | 1.98 | 0.98 | 2.10 | 1.04 | 2.32 | 1.22 |
| | .4 | 2.01 | 0.92 | 2.03 | 0.91 | 2.12 | 1.02 | 2.33 | 1.21 |
| 3.0 | .1 | 2.76 | 0.88 | 1.21 | 0.46 | 2.93 | 0.93 | 3.09 | 1.39 |
| | .2 | 2.89 | 0.85 | 2.74 | 0.89 | 3.08 | 0.91 | 3.10 | 1.39 |
| | .3 | 2.96 | 0.81 | 2.92 | 0.76 | 3.14 | 0.87 | 3.39 | 1.04 |
| | .4 | 3.00 | 0.74 | 2.97 | 0.72 | 3.16 | 0.84 | 3.42 | 1.01 |

Table 9
Means and Standard Deviations
from SIMIP on VC1PL Item Pool
Varying Acceptance Range

| Acceptance Range | | Ability Level | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| .1 | $\overline{X}_\theta$ | -1.938 | -1.713 | -0.994 | -0.491 | 1.101 | 1.873 | 2.913 |
| | $S_\theta$ | 0.430 | 0.794 | 0.573 | 0.810 | 0.976 | 0.676 | 0.447 |
| | Mni* | 3.64 | 5.56 | 6.84 | 8.52 | 11.12 | 9.72 | 6.24 |
| .2 | $\overline{X}_\theta$ | -2.747 | -2.133 | -1.193 | -0.152 | 1.208 | 2.268 | 2.889 |
| | $S_\theta$ | 0.790 | 0.520 | 0.779 | 0.544 | 0.739 | 0.686 | 0.540 |
| | Mni* | 6.96 | 8.88 | 12.56 | 14.44 | 15.44 | 10.56 | 7.20 |
| .3 | $\overline{X}_\theta$ | -2.955 | -2.085 | -1.311 | -0.021 | 1.026 | 2.229 | 3.109 |
| | $S_\theta$ | 0.823 | 0.555 | 0.943 | 0.385 | 0.578 | 0.581 | 0.510 |
| | Mni* | 7.00 | 10.00 | 11.24 | 16.96 | 17.24 | 12.96 | 7.68 |
| .4 | $\overline{X}_\theta$ | -3.171 | -2.404 | -1.346 | -0.007 | 0.869 | 2.234 | 2.950 |
| | $S_\theta$ | 0.690 | 0.538 | 0.681 | 0.344 | 0.399 | 0.775 | 0.579 |
| | Mni* | 7.08 | 8.08 | 14.60 | 18.44 | 19.72 | 14.64 | 9.64 |
| .5 | $\overline{X}_\theta$ | -3.157 | -2.242 | -1.051 | 0.160 | 0.941 | 2.340 | 3.117 |
| | $S_\theta$ | 0.606 | 0.791 | 0.619 | 0.755 | 0.546 | 0.780 | 0.497 |
| | Mni* | 8.16 | 11.40 | 17.04 | 18.64 | 19.28 | 14.32 | 9.48 |

Note. All runs made with 25 administrations per ability level, 20
item upper limit, and stepsize = .693.
*Mni = mean number of items administered

.2 to .4. In cases such as this, one should consider a combination of
the density of the item pool across the range of θ and whether a parti-
cular θ range should be estimated more precisely than others, in order
to decide on the best acceptance range value. Decisions regarding the
best value of program parameters cannot be made independent of consider-
ations such as the size and shape of the item pool to be used.

## Secondary Results

Secondary results include the comparison of the performance of actual
versus ideal item pools previously discussed. Table 11 shows this compari-
son, and overall, the ideal pool did not perform much better than the
ET1PL pool.

Another comparison was of the SIMIP and TREEIP programs on the same
pools using the same program parameter values. By looking at Table 2
and Table 6, one may see that the SIMIP did a reasonably good job of

Table 10
Means and Standard Deviations
from SIMIP on ET1PL Item Pool
Varying Acceptance Range

| Acceptance Range | | Ability Level | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| .1 | $\overline{X}_\theta$ | -2.528 | -2.200 | -1.174 | -0.111 | 0.974 | 2.001 | 3.299 |
| | $S_\theta$ | 0.559 | 0.667 | 0.700 | 0.569 | 0.903 | 0.471 | 0.781 |
| | Mni* | 6.64 | 9.24 | 17.16 | 26.60 | 27.80 | 16.44 | 11.16 |
| .2 | $\overline{X}_\theta$ | -2.989 | -2.159 | -1.144 | -0.016 | 0.926 | 2.152 | 3.451 |
| | $S_\theta$ | 0.491 | 0.559 | 0.731 | 0.332 | 0.362 | 0.464 | 0.765 |
| | Mni* | 7.20 | 14.52 | 22.40 | 31.60 | 33.60 | 22.36 | 13.40 |
| .3 | $\overline{X}_\theta$ | -3.103 | -2.475 | -1.162 | 0.003 | 1.016 | 2.161 | 3.024 |
| | $S_\theta$ | 0.576 | 0.594 | 0.630 | 0.239 | 0.401 | 0.410 | 0.747 |
| | Mni* | 7.60 | 12.40 | 27.96 | 36.72 | 37.88 | 25.32 | 18.16 |
| .4 | $\overline{X}_\theta$ | -3.064 | -2.359 | -1.121 | -0.094 | 1.043 | 2.073 | 3.054 |
| | $S_\theta$ | 0.615 | 0.315 | 0.582 | 0.261 | 0.316 | 0.336 | 0.520 |
| | Mni* | 10.20 | 18.00 | 31.40 | 39.00 | 39.36 | 31.52 | 20.12 |
| .5 | $\overline{X}_\theta$ | -3.378 | -2.465 | -1.088 | 0.031 | 0.993 | 1.920 | 3.195 |
| | $S_\theta$ | 0.716 | 0.715 | 0.510 | 0.394 | 0.356 | 0.389 | 0.584 |
| | Mni* | 10.24 | 18.48 | 35.08 | 39.80 | 39.48 | 35.12 | 20.76 |

Note. All runs made with 25 administrations per ability level, 40
item upper limit, and stepsize = .693.
*Mni = mean number of items administered

approximating the TREEIP results at $\theta = 2$ for the bimodal and skewed pools.
Also, from Table 6, it can be seen that increasing the number of tests
administered by the SIMIP did not dramatically change the means and stan-
dard deviations. Therefore, 25 administrations seemed adequate.

Finally, by comparing cells of Tables 7 and 10, one can see that
increasing the maximum number of items administered from 20 to 40 does
not substantially change the means and standard deviations from the SIMIP.
This comparison is not exact because the acceptance range of 0.25 used
for analyses in Table 7 does not precisely equal the value of 0.2 or 0.3
for acceptance range in Table 10. Neither is the stepsize of 0.7 in Table
7 exactly equal to 0.693 used in Table 10. However, the values seemed
close enough to make a comparison, and the result of this comparison seemed
to indicate that 20 items as an upper limit was adequate. Note that the
mean number of items recorded in both tables illustrated that the proce-
dure approached the upper limit in the middle range of $\theta$.

Table 11
Means and Standard Deviations from
SIM1P on ET1PL Item Pool and Comparable
Ideal Item Pool

| | | Ability Level | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| ET1PL | $\bar{X}_\theta$ | -3.061 | -2.175 | -1.026 | -0.065 | 1.029 | 1.950 | 2.913 |
| Pool | $S_\theta$ | 0.561 | 0.460 | 0.573 | 0.469 | 0.516 | 0.696 | 0.533 |
| Ideal | $\bar{X}_\theta$ | -3.036 | -2.404 | -1.037 | -0.017 | 1.148 | 2.222 | 3.070 |
| Pool | $S_\theta$ | 0.441 | 0.703 | 0.652 | 0.462 | 0.787 | 0.718 | 0.460 |

Note. All runs made with 25 administrations per ability level, 20
item upper limit, stepsize = 0.70, and acceptance range = 0.25.

## Discussion

It should be recalled that the basic emphasis of this study was to investigate the operational characteristics of a one-parameter tailored testing procedure when the item pool attributes (shape and size) and the program parameters (stepsize and acceptance range) were varied. In so doing, suggestions regarding the most preferred item pool and program parameter values were found based upon analyses of the tailored testing procedure's ability estimate bias and standard error at various points along the ability continuum. This strategy for investigating bias and standard error was motivated by the need to determine these values at several levels of $\theta$ across the continuum, since overall efforts on the project were directed toward developing a criterion-referenced tailored test. In criterion referenced testing, it is essential to identify effects of ability estimate bias and standard error on decisions made at several points along the ability scale. The research presented here, which was designed to determine optimal item pool attributes and program parameters to minimize bias and errors of measurement, provided a necessary foundation for further research and development of the criterion-referenced tailored testing strategy.

## Item Pool Shape

Overall conclusions about the most preferred item distribution were that the rectangular pool was most apt to yield the least bias and smallest standard error of ability estimates across the ability scale. One important caution was suggested by some of the results. When setting up an item pool for use with tailored testing procedures (especially those procedures having a parameter comparable to the acceptance range), it is important to look carefully at the frequency distribution of the item difficulties to be assured that no substantial gaps exist in any area of the continuum. Otherwise, one may expect poor estimation of ability at that region on

the continuum.  In this regard, one should view the estimates of true
ability +3.0 as understandably limited, in as much as the item pools did
not have any items beyond difficulty +3.0.  For best estimation of ability,
the pool should have a dense uniform distribution of items around the
ability level to be estimated.

## Item Pool Size

The methods employed for the investigation of the effects of item
pool size on the operation of the one-parameter maximum likelihood tailored
testing procedure were simulations, but theoretical methods have also been
proposed.  Lord (1970) suggested a formula for the number of items required
for a fixed stepsize procedure (selecting items more difficult by the
stepsize when correct responses were given and vice versa).  The formula
is

$$N = (1 + R/d) (n - R/2d) \qquad (2)$$

where +R is the range of item difficulties desired, d is the stepsize
and a submultiple of R, and n is the maximum number of items to be admin-
istered.  For example, if R were plus three to minus three, d were set
at 0.5, and n were twenty, the formula would give

$$119 = (1 + 3.0/0.5) (20 - 3.0/(2 \times 0.5)). \qquad (3)$$

With this set of values, 119 items would be required if the exact item
requested were to be available.

This formula does not directly apply to some tailored testing proce-
dures which use a variable rather than a fixed stepsize.  Also, most
testing procedures allow administration of slightly discrepant items from
those requested by the procedure (the acceptance range specified how dis-
crepant).  Procedures using a variable stepsize tend to require more items
because, as the procedures converge to an ability estimate, the stepsize
in effect becomes smaller and smaller.  Allowing items to be administered
which differ slightly from the requested item compensates to an extent
for the increase in number of items caused by the variable stepsize.
Another limitation of the formula is that several tailored testing procedures
administer items until a specified precision is reached instead of using
a preset maximum number of items as a stopping rule.

Another theoretical method of estimating how large an item pool should
be (Cliff, 1975) is to determine the number of items required to reach
a specified precision of ability estimation, given that equally spaced,
perfectly discriminating items are available.  With these ideal or optimal
circumstances, the precision of an ability estimate is equal to the difference
between adjacent items.  For example, an item pool with seven equally
spaced items from -3.0 to +3.0 would classify examinees into categories
1.0 scale unit apart.  The number of item responses required to make the
classification would be

$$K = \log_2 n \qquad (3)$$

where n is the size of the item pool, since $2^k$ is the number of branches in the tree diagram after k items are administered. By specifying the precision desired, e, the minimum item pool size can be determined by the range of ability, R, divided by e, plus one.

$$n = \frac{R}{e} + 1 \tag{4}$$

The minimum number of items administered to classify all ability levels in the tailored testing situation is

$$k = \log_2 [\frac{R}{e} + 1]. \tag{5}$$

Some results obtained by the application of the formulas based on the theoretical method for estimating the number of items needed in a pool, given the precision desired, have been indicated in Table 12. The requirements for pool size were computed for the range of ability, -3.5 to +3.5, given the desired classification interval size. As has been pointed out, these results are for a rectangular pool of hypothetical items with perfect discrimination and zero guessing probabilities. With these restrictions, the item pool sizes shown must be regarded as lower limits. The minimum session length indicates the fewest number of items that would have to be administered in order to classify an ability level within the capabilities of the item pool. These also are based on hypothetically perfect items and item pools, and should be considered as lower limits. The values in the column labelled simulated length are the number of items required to reach a best estimate using the most likely response pattern simulation. All results in this column are based on $\theta = 0.0$.

Table 12
Minimum Item Pool Requirements
for a Rectangular Idealized Pool Given
Classification Interval and Ability Range

| Ability Range | Classification Interval Size | Pool Size | Minimum Session Length | Simulated Length* |
|---|---|---|---|---|
| (-3.5, 3.5) | 0.5 | 15 | 3.9 | 2 |
| (-3.5, 3.5) | 0.25 | 29 | 4.9 | 4 |
| (-3.5, 3.5) | 0.125 | 55 | 5.8 | 8 |
| (-3.5, 3.5) | 0.0625 | 113 | 6.8 | 8 |
| (-3.5, 3.5) | 0.03125 | 225 | 7.8 | 7 |

*Note. Number of items administered to closest approximation of $\theta$ value within classification interval.

In some cases the simulated session length is less than the minimum predicted length because of the choice of ability level. Setting the

stepsize equal to 0.693 tends to keep the process near the middle of the
item pool, speeding up convergence for abilities near 0.0.  If an ability
of 3.0 had been used, the session length for classification interval .5
would have been 6, well over the minimum predicted values.  Thus, the mini-
mum session length refers to the number of items needed across the ability
range, and under specified circumstances fewer items may be required.

These results using simulated tests have been compared to actual
tailored testing convergence plots and found to be fairly good approxi-
mations (Reckase, 1976).  One observation of importance is that, from
convergence plots, it can be seen that giving too many easy items causes
bias in ability estimation.  Reckase (1975) has discussed this effect
in detail.

## Stepsize

The investigation of the stepsize program parameter suggests that
for tailored testing procedures using a fixed stepsize prior to having
correct and incorrect responses in the examinee's response string, a value
in the range of .5 to 1.0 is most apt to minimize ability estimate bias
and standard error.  To determine the precise stepsize value to use when
setting up a tailored testing procedure, one should look carefully at
the distribution of item difficulty of the particular item pool to be
used.  The testing procedure should select the first item from the middle
of the pool.  This item may not coincide with the most informative item
for $\theta = 0$, since the median difficulty for the pool may not equal 0.
The next step is to tentatively set the stepsize equal to 0.7 and deter-
mine whether items exist within the acceptance range at +1, +2, +3, and
+4 stepsizes away from the median difficulty item that the procedure
administered first.  The purpose here is to avoid setting the stepsize
at a value which will induce ability estimates during initial testing
which will "fall through" the item pool (i.e. premature termination of
testing when no items exist within plus or minus the acceptance range of
the ability estimate).  If the item difficulty distribution is uniformly
dense across the range of difficulty this will not pose much of a problem.

Another consideration when setting the stepsize value is to make
it small enough to assure that items exist within an acceptance range of
+4 stepsizes away from the median difficulty item in the pool.  This will
make the minimum number of items that would be administered equal to 5
for those who get all the items right or all the items wrong.  Depending
on the above considerations, the stepsize value may be set lower or higher
than the recommended 0.7.  As can be seen, the item pool size and diffi-
culty distribution, acceptance range, and stepsize interact in determining
the adequacy of the testing procedure.

The reason for including 0.693 as a potentially optimal stepsize in
this study was that when the first Rasch procedure, using raw ability, was
set up at the University of Missouri, a multiplicative stepsize equal to
2 was used with good results.  When the procedure was changed to operate
on log ability, an additive stepsize equal to $\log_e 2$ seemed promising.
This study suggests that indeed $\log_e 2 = 0.693$ was justifiably chosen for
the stepsize in the one-parameter tailored testing procedure.

## Acceptance Range

As has been indicated in the discussion of stepsize, setting the values of the program parameters (stepsize and acceptance range) should be performed in accord with the item pool attributes of the pool to be used for testing. If the item pool has a uniform density of item difficulties, one may set the acceptance range at a fairly low value (say 0.2). However, if "gaps" exist along the difficulty continuum, the acceptance range should be set large enough to avoid terminating the test due to a lack of any item within an acceptance range of the ability estimate. In general, an acceptance range equal to 0.3 appeared to satisfy the conditions of avoiding premature termination of testing and also minimizing bias induced by administering inappropriate items.

The program parameter denoted acceptance range is equivalent to specifying a minimum item information cutoff. Table 13 indicates the comparable item information cutoffs for the acceptance ranges investigated for this report. Many of the tailored testing systems presently in operation compute the item information for each item in the pool given the present ability estimate. For the one-parameter model, the information function is maximized when the difficulty of the selected item equals the ability estimate. For a discussion of information functions see Birnbaum (1968).

Table 13
Comparable Information Cutoffs
for Acceptance Range Values

| Acceptance Range | Information Cutoff |
|------------------|--------------------|
| .1               | .249               |
| .2               | .248               |
| .3               | .244               |
| .4               | .240               |
| .5               | .235               |

## TREE1P

A possible explanation for the larger standard deviation given by analyses run on the rectangular pool at the more extreme values of the ability continuum was suggested by a close look at the development of the propensity distribution by the TREE1P for the various shaped item pools.

A property of the TREE1P and the manner in which it developed the propensity distributions was that the standard deviation actually increased as more branches or levels resulted from items administered to more and more possible ability estimates. This increase of the standard deviation of ability estimates stabilized for the smaller item pools as the paths or branches of the "tree" terminated. For the larger pools (especially the 61 item pools), the standard deviation initially increased but as

branches were terminated the standard deviation came down. Figure 4 illustrates this property of the TREE1P when it was run on the 61 item rectangular pool with θ set equal to zero, stepsize equal to 0.693, and acceptance range equal to 0.30. This pattern of increasing standard deviation of ability estimates during the early formulation of the propensity distribution was evident for all shapes of the distributions of items in the pools.

However, the patterns of convergence to the final standard deviations yielded by the TREE1P were different for the various shapes of item pools at different ability levels. Tables 1, 2, and 3 show a general tendency for the standard deviations of ability estimates of the true abilities zero and one to be larger for the normal and bimodal pools than for the rectangular pools. But for ability levels two and three, the standard deviations of ability estimates were generally larger for the rectangular pools than for the normal and bimodal pools. This trend was consistent across most of the TREE1P analyses. The explanation proposed was that, because more items were available for administration to the more extreme levels of ability (i.e. θ = 2 and θ = 3) when the rectangular pool was used, the standard deviation of ability estimates was larger since the standard error was more accurately estimated. The standard deviations of the estimates from the normal and bimodal pools for these true ability levels were smaller, since paths or branches were often terminated because no items were available within the acceptance range of the estimated abilities. In short, when fewer items were in the pool around a particular true ability, there were fewer paths allowed to develop in the propensity distribution due to the stopping rules. Therefore, the standard deviation of ability estimates at that particular level was an underestimate. A logical check for this phenomenon was the prediction that when the acceptance range was made smaller, the drop in standard deviations for the more extreme ability levels would be more pronounced with the normal pool than for the rectangular. This did appear to be the case. The point is that the smaller standard deviations for ability levels 2 and 3 yielded by the TREE1P when normal or bimodal pools were used probably should not be weighted too heavily, as the tendency appears to be somewhat of an artifact of the procedure. The values obtained for the rectangular pools may well be more realistic.

## SIM1P

SIM1P was designed to score and administer items in the manner previously described based on the rationale that this approach was a reasonable simulation of the behavior of an examinee when interacting with a tailored test. The pseudo examinee with some specified true ability was presented an item of average difficulty from the pool, because, given we have no prior information about his ability, the best guess of an item appropriate for the examinee was one of average difficulty. Scoring of each item by determining the probability of a correct response using the examinee's θ in the one-parameter formula and then comparing this probability to a random number selected from a rectangular distribution between zero and one was deemed a reasonable simulation, assuming the one-parameter model was correct. Clearly, the larger the probability of a correct response

Figure 4
TREE1P Convergence to the
Standard Deviation of Theta
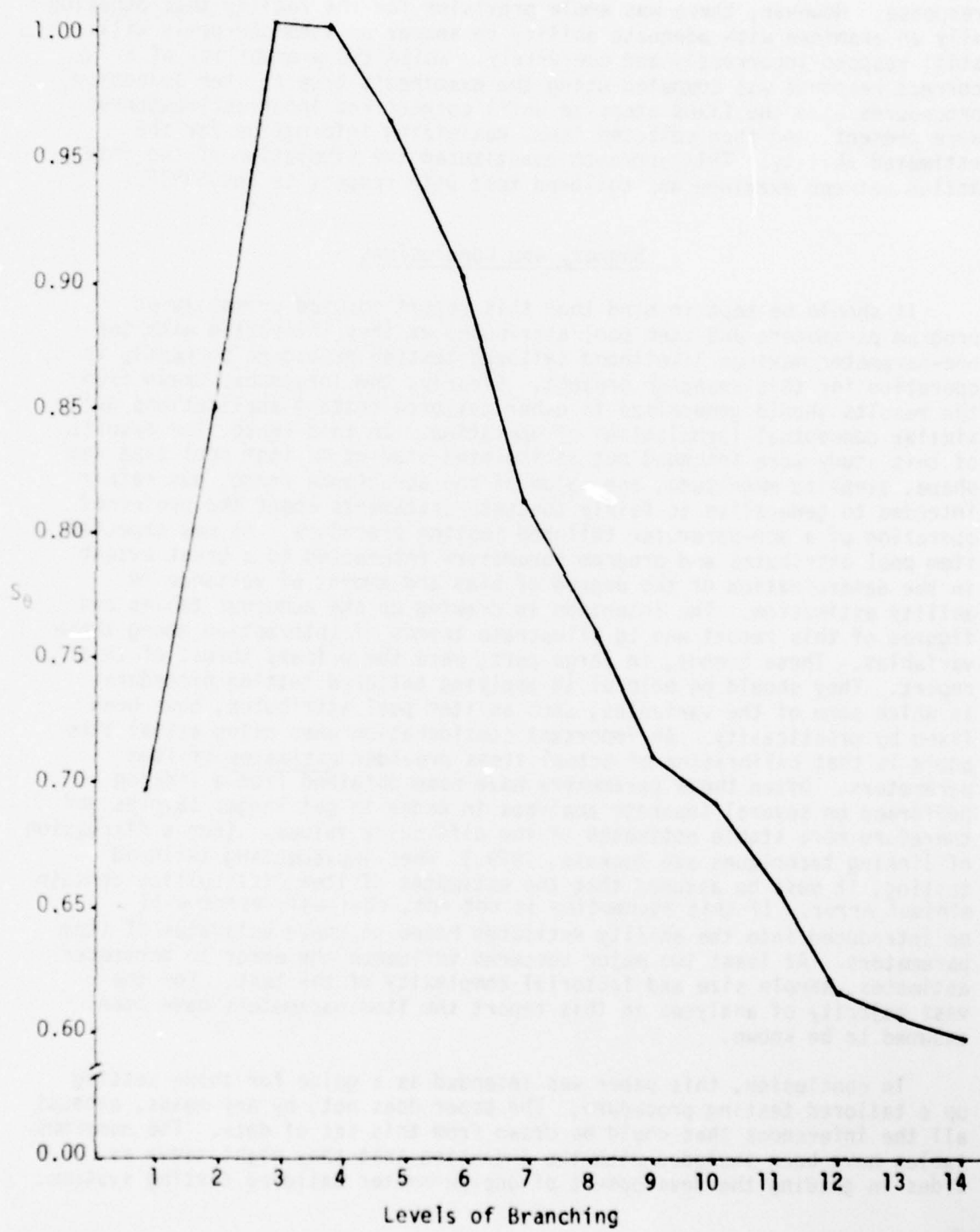
was, the greater the chance was that the random number generated was less than or equal to the probability specified by the model of a correct response. However, there was ample provision for the reality that occasionally an examinee with adequate ability to answer an item correctly will still respond incorrectly and conversely. While the probability of a correct response was computed using the examinee's true $\theta$, item selection procedures used the fixed stepsize until correct and incorrect responses were present, and then selected items maximizing information for the estimated ability. This approach constituted the simulation of the interaction between examinee and tailored test with respect to the SIMIP.

## Summary and Conclusions

It should be kept in mind that this report focused primarily on program parameters and item pool attributes as they interacted with the one-parameter maximum likelihood tailored testing procedure currently in operation for this research project. Clearly, the inferences drawn from the results should generalize to other tailored testing applications using similar conceptual formulations of operation. In this sense, the results of this study were intended not as isolated studies of item pool size and shape, stepsize magnitude, and value of the acceptance range, but rather intended to generalize to fairly concrete statements about the preferred operation of a one-parameter tailored testing procedure. As was expected, item pool attributes and program parameters interacted to a great extent in the determination of the degree of bias and amount of variance in ability estimation. The intention in drawing up the numerous tables and figures of this report was to illustrate trends of interaction among these variables. These trends, in large part, were the primary thrust of this report. They should be helpful in applying tailored testing procedures in which some of the variables, such as item pool attributes, have been fixed by practicality. An important consideration when using actual item pools is that calibration of actual items provides estimates of item parameters. Often these parameters have been obtained from a linking performed on several separate analyses in order to get larger samples and therefore more stable estimates of the difficulty values. (For a discussion of linking techniques see Reckase, 1979.) When implementing tailored testing, it must be assumed that the estimates of item difficulties contain minimal error. If this assumption is not met, obviously error will be introduced into the ability estimates based on these estimates of item parameters. At least two major concerns influence the error in parameter estimates, sample size and factorial complexity of the test. For the vast majority of analyses in this report the item parameters have been assumed to be known.

In conclusion, this paper was intended as a guide for those setting up a tailored testing procedure. The paper does not, by any means, exhaust all the inferences that could be drawn from this set of data. The numerous tables have been included with the intention that they might serve as aides in guiding the development of one-parameter tailored testing systems.

# REFERENCES

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Cliff, N. Personal Communication, Washington, D.C., June, 1975.

Jensema, C. J. Bayesian tailored testing and the influence of item bank characteristics. Paper presented at the Invitational Conference on Adaptive Testing, Washington, D.C., June, 1975.

Lord, F. M. Some test theory for tailored testing. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance, New York: Harper and Row, 1970.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Patience, W. M. Description of components in tailored testing. Behavior Research Methods and Instrumentation, 1977, 9(2), 153-157.

Pine, S. M., and Weiss, D. J. A comparison of the fairness of adaptive and conventional testing strategies (Research Report 78-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Reckase, M. D. An interactive computer program for tailored testing based on the one-parameter logistic model. Behavior Research Methods and Instrumentation, 1974, 6(2), 208-212.

Reckase, M. D. The effect of item choice on ability estimation when using a simple logistic tailored testing model. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., March, 1975. (ERIC Document Reproduction Service No. ED 106 342)

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

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

Weiss, D. J. Strategies of adaptive ability measurement. (Research Report 74-5), Department of Psychology, University of Minnesota, December, 1974.

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

Appendix A

Figure A-1
Frequency Distribution
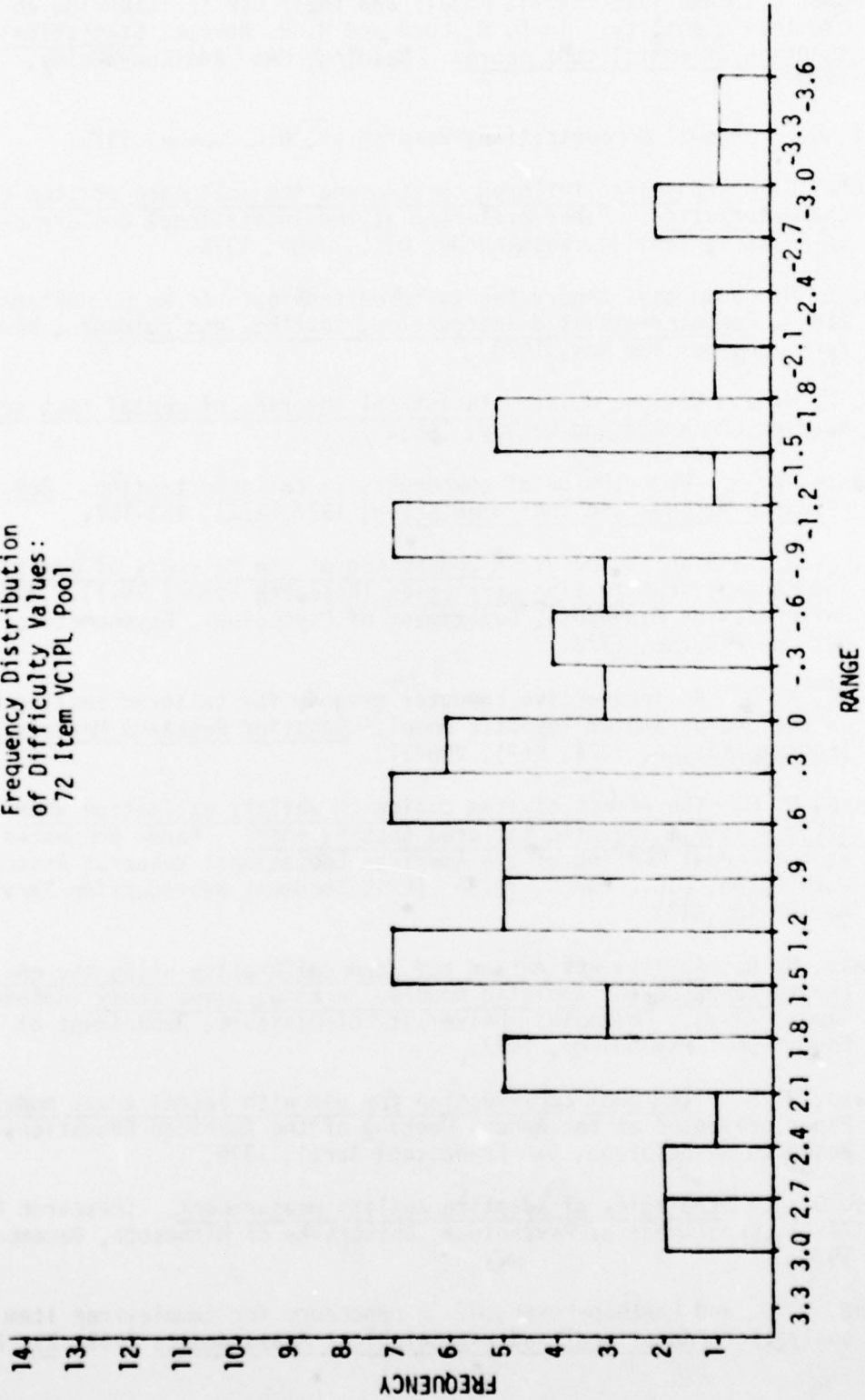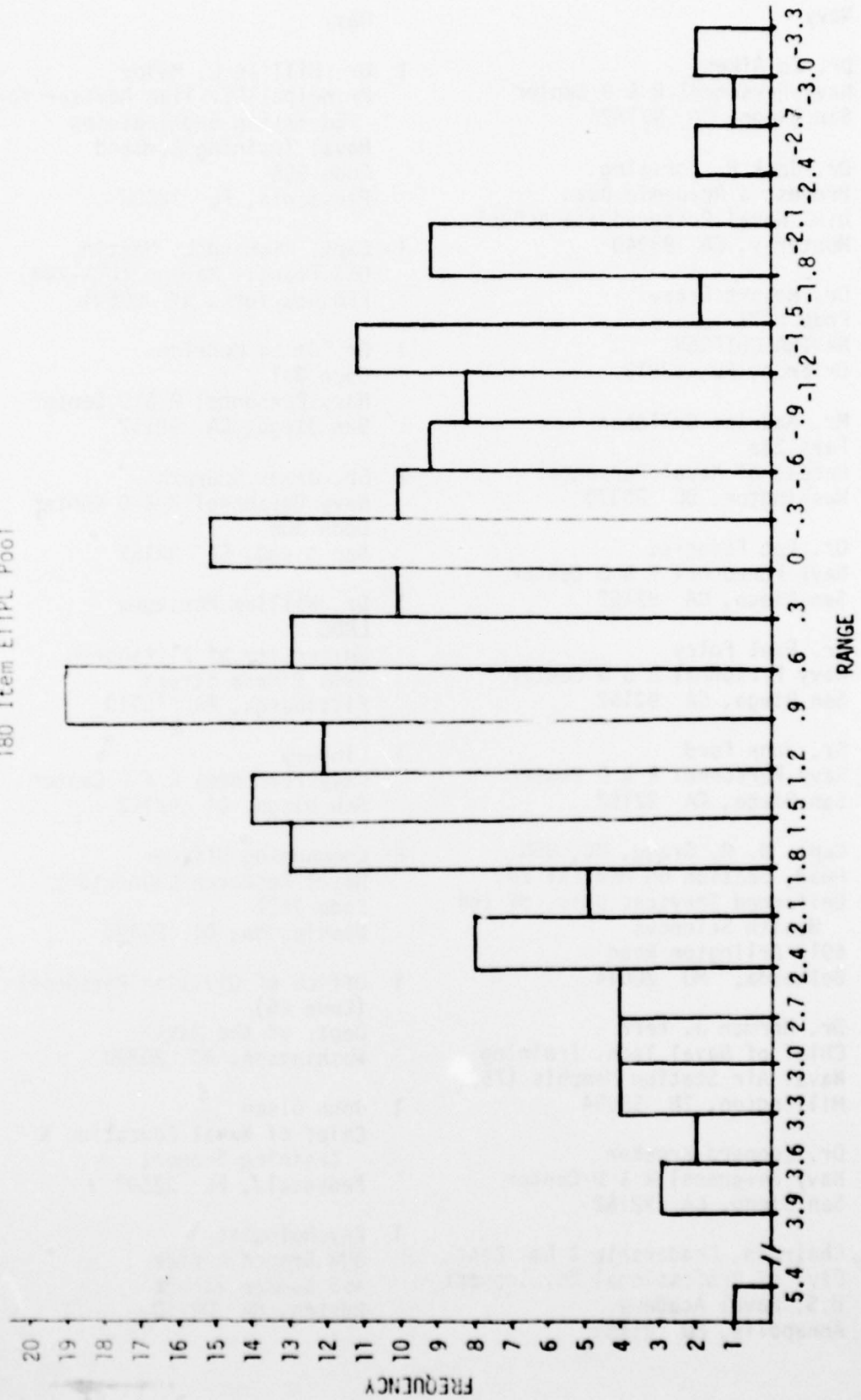of Difficulty Values:
72 Item VC1PL Pool



RANGE

FREQUENCY

Figure A-2
Frequency Distribution
of Difficulty Values:
180 Item ETIPL Pool



RANGE

FREQUENCY

# DISTRIBUTION LIST

### Navy

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

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

1   Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1   Mr. Maurice Callahan
Pers 23e
Bureau of Naval Personnel
Washington, DC 20370

1   Dr. Pat Federico
Navy Personnel R & D Center
San Diego, CA 92152

1   Dr. 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   Capt. D. M. Gragg, MC, USN
Head, Section on Medical Ed.
Uniformed Services Univ. of the
 Health Sciences
6917 Arlington Road
Bethesda, MD 20014

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

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

1   Chairman, Leadership & Law Dept.
Div. of Professional Development
U.S. Naval Academy
Annapolis, MD 21402

### Navy

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

1   Capt. Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1   Dr. James McBride
Code 301
Navy Personnel R & D Center
San Diego, CA 92152

2   Dr. James McGrath
Navy Personnel R & D Center
Code 306
San Diego, CA 92152

1   Dr. William Montague
LRDC
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

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

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

1   Office of Civilian Personnel
(Code 26)
Dept. of the Navy
Washington, DC 20390

1   John Olsen
Chief of Naval Education &
 Training Support
Pensacola, FL 32509

1   Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

Navy

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

1 Code 436
Office of Naval Research
Arlington, VA 22217

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
Office of Naval Research Branch
223 Old Marylebone Road
London, NW, 15th England

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

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval
Operations
Research, Development, & Studies
Branch (OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief
of Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 Dr. Richard A. Pollak
Academic Computing Center
U.S. Naval Academy
Annapolis, MD 21402

Navy

1 Mr. Arnold Rubenstein
Naval Personnel Support Tech.
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 A. A. Sjoholm
Tech. Support, Code 201
Navy Personnel R & D Center
San Diego, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Oper.
OP-987E
Washington, DC 20350

1 Dr. Alfred E. 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 CDR Charles J. Theisen, Jr. MSC,
USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Ronald Weitzman
Dept. of Administrative Sciences
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

Army

Air Force

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

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 Dr. Ralph Canter
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Ralph Dusek
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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

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

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

1 Director, Training Development
U.S. Army Administration Center
ATTN: Dr. Sherrill
Ft. Benjamin Harrison, IN 46218

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

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

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

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

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 Dr. G. A. Eckstrand
AFHRL/AS
Wright-Patterson AFB, OH 45433

1 CDR. Mercer
CNET Liaison Officer
AFHRL/Flying Training Div.
Williams AFB, AZ 85224

1 Dr. Ross L. Morgan (AGHRL/ASR)
Wright-Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DRMYP
Randolph AFB, TX 78148

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

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt., USAF
Program Manager
Life Scien-es Directorate
AFOSR
Bolling AFB, DC 20332

Air Force

1  Brian K. Waters, LCOL, USAF
   Air University
   Maxwell AFB
   Montgomery, AL  36112

Coast Guard

1  Mr. Richard Lanterman
   Psychological Research (G-P-1/62)
   U.S. Coast Guard HQ
   Washington, DC  20590

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

Marines

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

1  MCDEC
   Quantico Marine Corps Base
   Quantico, VA  22134

1  Dr. A. L. Slafkosky
   Scientific Advisor (Code RD-1)
   HQ, U.S. Marine Corps.
   Washington, DC  20380

Other DoD

12  Defense Documentation Center
    Cameron Station, Bldg. 5
    Alexandria, VA  22314
    ATTN:  TC

1  Dr. Dexter Fletcher
   Advanced Research Projects Agency
   1400 Wilson Blvd.
   Arlington, VA  22209

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

1  Major Wayne Sellman, USAF
   Office of the Assistant Secretary
       of Defense (MRA&L)
   3B930 The Pentagon
   Washington, DC  20301

Civil Govt.

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R & D Center
U.S. Civil Service Commission
1900 E. Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

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

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

1 Dr. Lalitha P. Sanathanan
Environmental Impact Studies Div.
Argonne National Laboratory
9700 S. Cass Avenue

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

Civil Govt.

1 Dr. Vern W. Urry
Personnel R & D Center
U.S. Civil Service Commission
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. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Erling B. Anderson
University of Copenhagen
Studiestraedt
Copenhagen
DENMARK

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

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

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

Non Govt.

1  Dr. Warner Birice
   Streitkraefteamt
   Rosenberg 5300
   Bonn, West Germany d-5300

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

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

1  Dr. David G. Bowers
   Institute for Social Research
   University of Michigan
   Ann Arbor, MI  48106

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 S. State St.
   Orem, UT  84057

1  Dr. John B. Carroll
   Psychometric Lab
   University of N. 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

Non Govt.

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

1  Dr. William Coffman
   Iowa Testing Programs
   University of Iowa
   Iowa City, IA  52242

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

1  Dr. Meredith Crawford
   Department of Engineering Adm.
   George Washington University
   Suite 805
   2101 L Street NW
   Washington, DC  20037

1  Dr. Hans Cronbag
   Education Research Center
   University of Leyden
   Boerhaavelaan 2
   Leyden
   The NETHERLANDS

1  Major I. N. Evonic
   Canadian Forces Pers. Applied
      Research
   1107 Avenue Road
   Toronto, Ontario
   CANADA

1  Dr. Leonard Feldt
   Lindquist Center for Measurement
   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

Non Govt.

1 Dr. Gerhardt Fischer
  Liebigasse 5
  Vienna 1010
  AUSTRIA

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

1 Dr. Edwin A. Fleishman
  Advanced Research Resources Org.
  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. Ross Greene
  CTB/McGraw Hill
  Del Monte Research Park
  Monterey, CA 93940

1 Dr. Alan Gross
  Center for Advanced Study
    in Education
  City University of New York
  New York, NY 10036

1 Dr. Ron Hambleton
  School of Education
  University of Massachusetts
  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

Non Govt.

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

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

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

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

1 Dr. Carl J. Jensema
  Gallaudet College
  Kendall Green
  Washington, DC 20002

1 Dr. Arnold F. Kanarick
  Honeywell, Inc.
  2600 Ridgeway Parkway
  Minneapolis, MN 55413

1 Dr. John A. Keats
  University of Newcastle
  Newcastle, New South Wales
  AUSTRALIA

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

1 LCOL. C. R. J. Lafleur
  Personnel Applied Research
  National Defense HQS
  101 Colonel by Drive
  Ottawa, CANADA K1A OK2

1 Dr. Michael Levine
  Department of Psychology
  University of Illinois
  Champaign, IL 61820

Non Govt.

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  08540

1  Dr. Scott Maxwell
   Department of Psychology
   University of Houston
   Houston, TX  77025

1  Dr. Sam Mayo
   Loyola University of Chicago
   Chicago, IL  60601

1  Dr. Allen Munro
   USC
   Behavioral Technology Labs
   3717 South Hope Street
   Los Angeles, CA  90007

1  Dr. Melvin R. Novick
   Iowa Testing Programs
   University of Iowa
   Iowa City, IA  52242

1  Dr. Jesse Orlansky
   Institute for Defense Analysis
   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. Steven M. Pine
   4950 Douglas Avenue
   Golden Valley, MN  55416

Non Govt.

1  Dr. Diane M. Ramsey-Klee
   R-K Research & System Design
   3947 Ridgmont Drive
   Malibu, CA  90265

1  Min. Ret. M. Rauch
   P II 4
   BUNDESMINISTERIUM DER VERTEIDIGUNG
   Postfach 161
   53 Bonn 1, GERMANY

1  Dr. Peter B. Read
   Social Science Research Council
   605 Third Avenue
   New York, NY  10016

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. Donald Rubin
   Educational Testing Service
   Princeton, NJ  08450

1  Dr. Larry Rudner
   Gallaudet College
   Kendall Greeen
   Washington, DC  20002

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

Non Govt.

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 Street
    Alexandria, VA 22314

1   Dr. Kazao Shigemasu
    University of Tohoku
    Department of Ed. Psych.
    Kawauchi, Sendai 982
    JAPAN

1   Dr. Edwin Shirkey
    Department of Psychology
    Florida Technological Univ.
    Orlando, FL 32816

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

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

1   Dr. Albert Stevens
    Bolt Beranek & Newman, Inc.
    50 Moulton Street
    Cambridge, MA 02138

1   Dr. Patrick Suppes
    Institue for Mathematical Studies
      in the Social Sciences
    Stanford University
    Stanford, CA 94305

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

Non Govt.

1   Dr. Brad Sympson
    Elliott Hall
    University of Minnesota
    75 E. River Road
    Minneapolis, MN 55455

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

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

1   Dr. Robert Tsutakawa
    Dept. of Statistics
    University of Missouri
    Columbia, MO 65211

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

1   Dr. Howard Wainer
    Bureau of Social Science Research
    1990 M. Street, NW
    Washington, DC 20036

1   Dr. Thomas Wallsten
    Psychometric Laboratory
    Davie Hall 013A
    University of North Carolina
    Chapel Hill, NC 27514

1   Dr. John Wannous
    Department of Management
    Michigan University
    East Lansing, MI 48824

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

Non Govt.

1 Dr. Susan Whitely
Psychology Department
University of Kansas
Lawrence, KA  66044

1 Dr. Wolfgang Wildgrube
Streitkraefteamt
Rosenberg 5300
Bonn, WEST GERMANY D-5300

1 Dr. Robert Woud
School Examination Department
University of London
66-72 Gower Street
London WCIE 6EE
ENGLAND

1 Dr. Karl Zinn
Center for Research on
   Learning & Teaching
University of Michigan
Ann Arbor, MI  48104