DOCUMENT RESUME

ED 208 006 TM 810 686

AUTHOR	Reckase, Mark D.
TITLE	The Use of the Sequential Probability Ratio Test in Making Grade Classifications in Conjunction with Tailored Testing.
INSTITUTION	Missouri Univ., Columbia. Tailored Testing Research Lab.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO	ONR-RR-81-4
PUB DATE	Aug 81
CONTRACT	N00014-77-C-0097
NOTE	27p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Classification; Comparative Analysis; *Computer Assisted Testing; Cutting Scores; Decision Making; Higher Education; *Latent Trait Theory; Maximum Likelihood Statistics; *Test Reliability
IDENTIFIERS	One Parameter Model; *Sequential Probability Ratio Test (Wald); *Tailored Testing; Test Length; Three Parameter Model

ABSTRACT

This report describes a study comparing the classification results obtained from a one-parameter and three-parameter logistic based tailored testing procedure used in conjunction with Wald's sequential probability ratio test (SPRT). Eighty-eight college students were classified into four grade categories using achievement test results obtained from tailored testing procedures based on maximum information item selection and maximum likelihood ability estimation. Tests were terminated using the SPRT procedure. The results of the study showed that the three-parameter logistic based procedure had higher decision consistency than the one-parameter based procedure when classifications were repeated after one week. Both procedures required fewer items for classification into grade categories than a traditional test over the same material. The three-parameter procedure required the fewest items of all, using an average of 12 to 13 items to assign a grade. (Author)

THE USE OF THE SEQUENTIAL PROBABILITY RATIO TEST IN MAKING GRADE CLASSIFICATIONS IN CONJUNCTION WITH TAILORED TESTING

Mark D. Reckase

Research Report 81-4
August 81

Tailored Testing Research Laboratory
Educational Psychology Department
University of Missouri
Columbia, MO 65201

Prepared under Contract No. N00014-77-C-0097, NR150-395
With the Personnel and Training Programs
Psychological Sciences Division
Office of Naval Research

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 81-4	2. GOVT ACCESSION NO	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Use of the Sequential Probability Ratio Test in Making Grade Classifications in Conjunction with Tailored Testing		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Mark D. Reckase		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0097
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Educational Psychology University of Missouri Columbia, MO 65211		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N Proj.:RR042-04 T.A.:042-04-01 W.U.:NR150-395
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE August 1981
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

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)

Tailored Testing
Computerized Adaptive Testing
Sequential Probability Ratio Test
One-Parameter Model
Three-Parameter Model
Decision Making

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

This report describes a study comparing the classification results obtained from a one-parameter and three-parameter logistic based tailored testing procedure used in conjunction with Wald's sequential probability ratio test (SPRT). Eighty-eight college students were classified into four grade categories using achievement test results obtained from tailored testing procedures based on maximum information item selection and maximum likelihood ability estimation. Tests were terminated using the SPRT procedure.

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

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#20 (Continued)

The results of the study showed that the three-parameter logistic based procedure had higher decision consistency than the one-parameter based procedure when classifications were repeated after one week. Both procedures required fewer items for classification into grade categories than a traditional test over the same material. The three-parameter procedure required the fewest items of all, using an average of 12 to 13 items to assign a grade.

CONTENTS

THE USE OF THE SEQUENTIAL PROBABILITY RATIO TEST IN MAKING GRADE CLASSIFICATIONS IN CONJUNCTION WITH TAILORED TESTING

In many testing applications, the major use of the obtained score is to classify a person as being above or below some criterion score. Examples of such uses of test results include the screening of job applicants and the classification of students as masters and non-masters when using the mastery learning paradigm (Bloom, 1971). For such applications it is not necessarily required that the person's ability be accurately estimated, but only that the measurements be sufficiently precise that the examinees can be accurately classified.

When making such classifications, the accuracy of measurement required in making the decision is dependent upon how far from the cutting score the person is located. If the examinee is far above or below the cutting score, minimal accuracy will be required. If the examinee is close to the cutting score, high precision will be required. Since the accuracy of an ability estimate is dependent to a large extent on test length, it follows that shorter tests can be used if a person's ability were a substantial distance from the cutting score. Depending on the number of individuals who are far from the cutting score, the average length of test needed for classification might be substantially reduced over what is commonly used.

Based on this analysis, an optimal procedure for testing examinees for classification purposes would be to check the accuracy of classification after each item is administered. If the accuracy were sufficiently high, testing could stop. If the accuracy were not high enough, another item would be administered.

Exactly this type of procedure was developed by Wald (1947) to assist in quality control work during World War II. His procedure was designed to determine whether a batch of parts was acceptable based on whether it contained a sufficiently low number of defectives. The basic concept behind the procedure is to take an observation from the batch and determine the probability of the observation under the hypothesis of an acceptable or unacceptable batch. A ratio is formed by dividing the probability of the observation coming from an acceptable batch by the probability of it coming from an unacceptable batch. If the ratio is sufficiently large, the batch is considered acceptable and if it is sufficiently small, the batch is considered unacceptable. If the ratio is near 1.0, another observation is randomly selected. A new ratio is then formed using all of the previous observations. The process continues until a decision is reached. Because of the sequential nature of the process, it has been labeled the Sequential Probability Ratio Test (SPRT).

Since its development, the SPRT has been widely used for quality control work (Govindarajulu, 1975). However, only recently has it appeared in the mental testing literature. Ferguson (1970) used the SPRT procedure to determine whether 75 students had mastered material in a hierarchically arranged set of instructional units. His procedure randomly generated items by computer using item forms and then administered the items using a computer terminal. He found a substantial reduction in testing time and in the number of items required to make a decision. The procedure was found to be in 99% agreement with the longer tests traditionally used to make the decisions.

No other studies were found that actually made real time decisions using the SPRT procedure. However, Epstein & Knerr (1978) did present the results of a real data simulation using Army proficiency testing response data. They found that only 33% as many items were needed for the SPRT based procedure without loss in decision accuracy. Sixtl (1974), Kalish (1980), and Kingsbury and Weiss (1980) present the results of simulation studies showing that the SPRT procedures result in a substantial reduction in the number of items required to make decisions. Thus, all the research to date supports the contention that SPRT based procedures lead to increased testing efficiency.

Despite the promising results reported in the studies listed above, none of the procedures described take full advantage of the quality items in the item pool. That is, by randomly selecting items, the best items for making the classification decision may not be administered. A better procedure would be to select the items from the item pool that would be most informative for making the decision using a tailored testing paradigm. Reckase (1978) has shown that such a procedure could be used with the SPRT as long as local independence could be assumed. In a series of simulation studies (Reckase, 1980a, 1980b), he demonstrated that SPRT procedures will work with tailored testing. Further, a three-parameter logistic based procedure was found to give better results than a one-parameter logistic based procedure.

With the positive results obtained at this time it seems prudent to evaluate the quality of SPRT/tailored testing procedures for actual decisions. The purpose of this report is to present some results of the operation of the SPRT/tailored testing hybrid in the context of grade classification. Further, one-parameter and three-parameter logistic model based procedures will be compared on the basis of decision consistency. The overall criterion for success will be a comparison with traditional grading procedures.

The SPRT Procedure

The SPRT procedure has been described in detail elsewhere (Wald, 1947; Epstein & Knerr, 1978; Reckase, 1980a) so only a brief description will be given here. The basic equations will be presented along with the procedures for describing the characteristics of the decision making process.

As described above, the basic philosophy behind the SPRT procedure is to determine the probability of the observed responses for two alternative hypotheses and then form the ratio of the probabilities. A large ratio favors one of the hypotheses and a small ratio favors the other. For example, if H_1 is the hypothesis that the ability (θ) for a person is equal to θ_1, and H_2 is the hypothesis that the ability equals θ_2, the probability of the obtained responses, $x_1, x_2, \ldots, x_n$, given these hypotheses would be:

$$P(x_1, x_2, \ldots, x_n \mid \theta_1) = \prod_{i=1}^{n} P(x_i \mid \theta_1) \quad (1)$$

and

$$P(x_1, x_2, \ldots, x_n \mid \theta_2) = \prod_{i=1}^{n} P(x_i \mid \theta_2) \quad (2)$$

under the local independence assumption of latent trait theory. The values of $P(x_i|\theta_2)$ would be computed using the appropriate latent trait model assuming known item parameters from a previous item calibration. Assuming $\theta_1 < \theta_2$, the probability ratio would then be formed as

$$L = \frac{P(x_1, x_2, \ldots, x_n|\theta_1)}{P(x_1, x_2, \ldots, x_n|\theta_2)} \tag{3}$$

If this ratio were sufficiently large H_2 would be rejected, and if the ratio were sufficiently small H_1 would be rejected. The determination of what constitutes large and small depends upon the error rates that are considered acceptable.

Suppose α is the probability of accepting H_1 when H_2 is really true and β is the probability of accepting H_2 when H_1 is really true. Wald (1947) has shown that a good approximation to the decision points needed for the probability ratio (Equation 3) can be obtained by the following two expressions:

$$\text{Upper decision point} = A = \frac{1-\beta}{\alpha} \tag{4}$$

and

$$\text{Lower decision point} = B = \frac{\beta}{1-\alpha} \tag{5}$$

Thus, if Equation 3 gives a result larger than A, H_1 should be accepted with an error rate of approximately α, and if the expression yields a value less than B, H_2 should be accepted with an error rate of approximately β.

The procedure described above assumes that a decision is to be made between two simple hypotheses: $H_1{:}\theta=\theta_1$ or $H_2{:}\theta=\theta_2$. Wald (1947) has generalized this procedure to making decisions concerning complex hypotheses such as $H_0{:}\theta<\theta_c$ and $H_1{:}\theta\geq\theta_c$. This is a much more useful set of hypotheses because it matches the decision process used in making classifications above or below a criterion score.

In order to test a complex hypothesis using the SPRT, an indifference region must first be specified around the cutting score, θ_c, for the decision. The indifference region is the area around the cutting score in which either classification is considered equally good. For example, if θ_c is the cutting score for making the decision, persons sufficiently close to θ_c could be classified either high or low without appreciable loss. Sufficiently close is defined here as being between θ_1 and θ_2 when $\theta_1>\theta_c>\theta_2$. If a person were outside the region from θ_1 to θ_2 and were misclassified, the error would be considered serious.

The use of the SPRT to test complex hypotheses works the same as for the simple hypotheses except that the limits of the indifference region are used in Equation 3 to form the probability ratio instead of the hypothesized true values. The upper and lower decision points for the test are determined in exactly the same way as before (Equations 4 and 5). However, now the operation of the SPRT is controlled not only by the α and β error rates, but also by the width of the indifference region. The higher the error rates and the wider the indifference region, the fewer the items that need to be administered.

The quality of operation of the SPRT procedure is usually judged on the basis of two mathematical functions called the operating characteristic (OC) function and the average sample number (ASN) function. The OC function is defined as

$$OC(\theta) = P(\text{classified below } \theta_c | \theta).$$

This function should have values close to 1.0 for $\theta<\theta_c$ and values close to 0.0 for $\theta>\theta_c$. To the extent that this function drops quickly from a value near 1.0 to near 0.0 in the indifference region, the SPRT procedure is working well.

The ASN function is defined as the average number of observations needed to make a decision as a function of θ. This function is typically peaked, with high values near the cutting score and decreasing values with increased distance from the cutting score. Both the OC function and the ASN function are dependent on the size of the error rates and the width of the indifference region. A narrow indifference region and/or low error rates result in a steep OC function and require a large number of observations for decisions. High error rates and/or a wide indifference region flatten the OC function and reduce the number of observations required. Thus, the price paid for high precision is a greater number of observations. More detailed information concerning the OC and ASN functions can be found in Wald (1947), Reckase (1980a), or Epstein and Knerr (1978).

Tailored Testing Procedure

Tailored testing procedures are defined by their methods of item selection and ability estimation. The procedure used in this study selects items to maximize the value of the information function (Birnbaum, 1968) at the previous ability estimate. Ability was estimated using an empirical maximum likelihood approach. The procedure is described in detail by McKinley & Reckase (1980), so it will not be described again here. The above tailored testing procedure was used with both the one-parameter logistic (1PL) and the three-parameter logistic (3PL) models in the study reported here.

Tailored Testing/SPRT Hybrid

The procedure used to administer the test items in this study used components of both tailored testing methodology and the SPRT. Items to be administered in the process of the computerized test were selected using the maximum information criterion (Birnbaum, 1968; McKinley & Reckase, 1980). After the response to each item was obtained, the value of the probability ratio (Equation 3) was computed and a decision was made to classify high, classify low, or to administer another item. If another item were to be administered, a maximum likelihood ability estimate was obtained and a new item was selected to maximize the information function at that ability estimate and administered to the examinee. The process continued until a classification decision had been made or until 20 items had been administered. After 20 items, ratios above 1.0 resulted in a high classification, and ratios below 1.0 resulted in low classification.

Research Design

The purpose of the research reported here was to compare 1PL and 3PL based procedures for making classification decisions using the SPRT. Since the true classifications were unknown, a consistency of classification design was used as a criterion for evaluation. To facilitate the comparison of decision consistency a test-retest design was used in which tailored tests based on both the 1PL and 3PL models were administered to the same individuals in two sessions one week apart. In the first session the 1PL and 3PL tailored tests were administered as described above without a break in between. From the student's point of view, only one test was administered. In the second session, the same procedure was followed, only the order of presentation of the 1PL and 3PL procedures was reversed to counterbalance fatigue effects. The initial order of presentation of the 1PL and 3PL procedures was randomly assigned to the students.

Within the tailored tests, three grade placement decisions were made using the SPRT procedure. Based on the test information, students were placed above or below the A/B grade cutoff, the B/C grade cutoff, and the C/D grade cutoff. Thus, if a student were classified below the A/B cutoff, and above the B/C cutoff, a grade of B would be assigned. The grade cutoffs for the study were set to be consistent with those used on the traditional test using the test characteristic curve.

Before the cutoffs could be set, the traditional test first had to be linked to the tailored testing item pool. This was done so that the cutoffs determined from the traditional test would be on the same scale as the tailored test ability estimates. The linking was performed using the major axis method for the 1PL model, and the maximum likelihood method for the 3PL model. See Reckase (1979a) for a more detailed description of these procedures.

The traditional test used as a basis for the grade cutoffs was a 50 item multiple choice test over the area of classroom evaluation procedures. The test and the population of students who took part in the study were from an introductory course on educational measurement techniques. The grade classification region for the traditional test in terms of raw scores were: 42-50, A; 33-41, B; 29-32, C; and 28 and below, D. Based on these score ranges, the A/B cutoff was set at 41½, the B/C cutoff at 32½, and the C/D cutoff at 28½. The 1PL ability scale cutoffs corresponding to the raw score cutoffs were A/B, 2.24; B/C, .95; and C/D, .46. The cutoffs on the 3PL ability scale were: A/B, .78; B/C, -.85; and C/D, -1.39. These values were determined by finding the points in the latent trait scales that were equivalent to the raw score points.

Along with the cutting points, an indifference region and the α and β error rates were needed to totally specify the SPRT procedure. A reasonable indifference region for the test was thought to be one standard error of measurement on either side of the cutting point. Based on the traditional test reliability of .60 for the sample of students used in the study, the standard error of measurement in 1PL and 3PL ability units was .45. Thus, the indifference regions were set at A/B, 2.69 to 1.79; B/C, 1.40 to .50; and C/D, .91 to .01 for the 1PL procedure and A/B, .23 to 1.33; B/C, -1.30 to -.40; and C/D, -1.84 to -.94 for the 3PL procedure. The differences in indifference regions for the two procedures were due to differences in the way the origins of the ability scales were defined.

Since it was considered a more serious error to classify someone high incorrectly than low incorrectly, α was set at .02 and β was set at .10. Using Equations 4 and 5, the decision points for the SPRT were computed to be A=45 and B=.102. This resulted in a classification in the higher grade category if Equation 3 resulted in a value greater than 45, in the lower grade category if the value was below .102, and continued testing if the result was between 45 and .102. The same A and B values were used for both the 1PL and 3PL procedures.

The sample used in this study consisted of 88 student volunteers from an undergraduate introductory measurement course. Of the 88 students, 21 were male and 67 female. The group consisted of 19 juniors, 67 seniors, and 2 graduate students. The tailored tests were administered the week following a classroom test over the same content. The examinees were told that the tailored test score would be substituted for the classroom test score if they performed better on the tailored test, and that they would receive extra credit points for completing the requirements of the study.

Analyses

The major analysis performed in this study was the comparison of the grade classifications over the test-retest period. This analysis was to show which procedure (1PL or 3PL) gave more consistent grade classification over the one week time period. Since the grade scale yields mainly categorical results, a phi coefficient derived from the chi-square contingency table was used for this analysis. The same analysis was also performed to determine which procedure made grade classifications that were more similar to those obtained from a traditional classroom test.

Along with the above analyses, the distributions of grades for the two procedures were determined and compared. The number of items required for a decision were also tabulated for each procedure and the mean number of items required were compared using a two-way ANOVA. Session and procedure were the independent variables in this analysis, with repeated measures over both session and procedure.

Results

The direct result of the tailored testing procedure in this study is the classification of students into grade categories using the SPRT paradigm. The results of this grade classification for the 1PL and 3PL tailored testing procedure, and the traditional classroom test are shown in Table 1. This table presents the frequency distribution of the grades for each procedure and each testing session. The means and standard deviations are also presented to summarize the distributions even though the data are only ordinal.

From these results, a tendency can be seen for the 1PL procedure to grade slightly easier than the 3PL procedure. The traditional test assigned the highest average grade of all the procedures. This can probably be explained by the fact that the classroom test was the test studied for and it was taken first. The standard deviations of grades for the 1PL and 3PL procedures were about the same, with a slight increase in the second testing session. The traditional test had the smallest standard deviation of all of the procedures.

Table 1

Grade Distributions for the 1PL and 3PL Tailored Tests and the Traditional Classroom Test

Session	Grade	Procedure 1PL	3PL	Traditional
	A(4)	13	6	8
1	B(3)	60 $\bar{x}$=2.78	58 $\bar{x}$=2.59	78 $\bar{x}$=2.91
	C(2)	20 s.d.=.75	26 s.d.=.75	10 s.d.=.56
	D(1)	7	10	4
	A(4)	18	12	
2	B(3)	54 $\bar{x}$=2.78	50 $\bar{x}$=2.65	
	C(2)	17 s.d.=.88	27 s.d.=.83	
	D(1)	11	10	

Note: The values presented in the table are percentages of 88 cases.

The results of the consistency of classification analysis are presented in Table 2 along with a comparison with the grades assigned by the traditional classroom exam over the same course content and the final grade in the course. As can be seen from this table, the consistency of the 3PL/SPRT procedure was substantially higher than the 1PL/SPRT procedure (phi = .938 vs. .662; t = 5.19, $p < .01$).

Table 2

Phi Coefficients Showing the Consistency of Grade Classifications and the Relationship With Traditional Grading Practices

Test	Test 1PL-1	1PL-2	3PL-1	3PL-2	Course Exam	Final Grade
1PL-1		.662	.340	.489	.486	.679
1PL-2			.448	.645	.495	.710
3PL-1				.938	.376	.461
3PL-2					.490	.649

Note: All phi coefficients are based on 88 cases.

The relationship between the tailored testing results and the traditional grading schemes show a more confusing pattern. The 1PL procedure had a correlation of around .5 with the exam grades and about .7 with the final grades. This was unexpected because the course exam was on the same material as the tailored test, while the final grade was based on a composite of three exams over different content areas. The correlations of the 3PL procedure with the course grade gave a similar pattern of results, but the grades assigned by the first 3PL session had lower phi coefficients. The results from the second testing were about the same magnitude as the 1PL results.

The data on the mean number of test items required to make the grade classifications are presented in Table 3. Since the tailored testing procedures were terminated if a grade decision were not made at or before 20 items, the table also gives the percent of cases making classifications in 20 items or less. As can be seen from this table, the 1PL procedure seldom was able to make classification decisions in 20 items or less, while about half the time the 3PL procedure could. Overall, the 3PL procedure required significantly fewer items to make a decision than the 1PL procedure ($\bar{x}$=13.41 vs. 18.14). Significantly fewer items were also required for the second testing session. The ANOVA on the number of items required for classification is given in Table 4. The low number of items required for a grade classification is even more dramatic when compared to the 50 items used to make the grade classifications with the traditional test.

Table 3

Average Number of Items Required
To Make Grade Classifications
by Procedure and Session

	Procedure			
	1PL		3PL	
Session	1	2	1	2
Percent using 20 items or less	5.70	6.80	50.00	53.40
$\bar{x}$ for cases 20 items or less	11.20	14.50	9.02	11.80
$\bar{x}$ for all cases (N=88)	18.61	17.66	13.97	1[illegible].35
S.D. for all cases	2.85	4.00	4.94	5.00

Table 4

ANOVA Results on Number of Items Administered With Model and Session as Independent Variables and Repeated Measures on Both Variables

Source	SS	df	MS	F	p
Model	1966.55	1	1966.55	96.55	.00
Session	94.10	1	94.10	6.59	.01
Model x Session	.56	1	.56	.03	.85
Error (model)	1771.95	87	20.37		
Error (session)	1242.40	87	14.28		
Error (interaction)	1397.94	87	16.07		

Discussion

The major thesis of this paper is that the number of items required to make a decision concerning the classification of individuals above or below a cutting score can be substantially reduced from the number traditionally used. This can be done because abilities far removed from the cutting score need not be measured as precisely as those who are near the cutting score. In order to implement a testing procedure that can modify the length of the test as a function of the examinee's ability, a tailored testing procedure based on maximum information item selection and maximum likelihood ability estimation (McKinley and Reckase, 1980) was combined with Wald's (1947) Sequential Probability Ratio Test.

Common wisdom in test theory indicates that in order to accurately classify individuals into two groups, the items should be selected to be most informative at the cutting score (Lord & Novick, 1968). This could be done in this situation by selecting items with maximum information at the cutting score and using the usual SPRT procedure. However, in this case three cutting scores were present (A/B, B/C, C/D) so the usual tailored testing item selection procedure of choosing items to give maximum information at the most recent ability estimate was used.

Beyond demonstrating the economics of the tailored testing/SPRT hybrid over traditional testing, the purpose of this paper was to compare tailored tests based on the 1PL model with tailored tests based on the 3PL model. The results showed that the 3PL procedure is clearly more consistent than the 1PL procedure, but that the relationship to the grades based on the classroom tests was about the same or a little worse for the 3PL procedure. This may be explained by the fact that the 1PL model tends to give ability estimates that are the sum of the components in a test while the 3PL based tests tend to give ability estimates that are more pure measures of the first principal component of a test (see

Reckase, 1979, for a more thorough discussion). The larger correlations with the final grades than with the exam grades is probably due to the higher reliability of the final composite based on the sum of three exams. The generally low correlations with the course grades were probably due to the low reliability of the course exams (.60) and differences in method variance.

The test length analysis resulted in several interesting findings. First, the 1PL based procedure had great difficulty in classifying students into grade categories with less than 20 items. The three parameter procedure could make the classification with less than 20 items about half the time. On the average, the 3PL procedure required about 5 items less for classification than the 1PL procedure. This shorter test length with higher consistency of classification is probably a result of the advantage obtained by using the item discrimination parameter in item selection. Since the 1PL procedure assumes that all items are of equal discriminating power, only the nearness of the item difficulty parameter to the most recent ability estimate affects item selection. In selecting items using maximum information with the 3PL procedure, discrimination, guessing, and difficulty parameters contribute to selection. This results in the administration of higher quality items overall. The fewer test items required in the second session may be due to greater familiarity with the testing system resulting in fewer mistakes in using the terminals. McKinley & Reckase (1980) give more details concerning the characteristics of the items actually administered in this study.

Summary and Conclusions

The purpose of this paper has been to compare two tailored testing based decision making procedures using the Sequential Probability Ratio Test. The procedures were based on the one-parameter logistic model and the three-parameter logistic model. The procedures were also compared to traditional paper and pencil test based grades.

The results of the study showed that the 3PL based tailored test/SPRT procedure had higher decision consistency and required fewer test items than the 1PL based procedure. The tailored testing/SPRT procedure also required substantially fewer items than the traditional classroom test ($\bar{x}$=13.4 vs. 50). These results indicate that a substantial increase in efficiency can be obtained through the use of tailored testing/SPRT procedures, but that the grades assigned may not be the same as those given using a traditional method. Of the two procedures used in this study, the 3PL based method was superior to the 1PL method in decision consistency and number of items required. Both procedures had about the same correlations with the traditional grades.

References

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

Bloom, B. S. Mastery learning. In J. H. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart and Winston, 1971.

Epstein, K. I. & Knerr, C.S. Application of sequential testing procedures to performance testing. In D. J. Weiss (Ed.), Proceedings of the 1977 computerized adaptive testing conference. Minneapolis: University of Minnesota. 1978.

Ferguson, R. A model for computer-assisted criterion-referenced measurement. Education, 1970, 91, 25-31.

Govindarajulu, Z. Sequential statistical procedures. New York: Academic Press, 1975.

Kalisch, S. J. A model for computerized adaptive testing related to instructional situations. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota, 1980.

Kingsbury, G. G. & Weiss, D. J. A comparison of ICC-based adaptive mastery testing and the Waldian probability ratio method. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota, 1980.

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

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

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

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

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

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

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

Sixtl, F. Statistical foundations for a fully automated examiner. Zeitschrift fur Entwichlungspsychologie und Padagogische Psychologie, 1974, 6, 28-38.

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

Navy

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

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

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

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

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

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

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

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

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

Navy

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

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

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

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

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

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

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

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

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

Navy

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

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

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

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

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

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

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

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

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

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

Navy

1 Dr. Worth Scanland, Director
Research, Development, Test & Evaluation
N-5
Naval Education and Training Command
NAS, Pensacola, FL 32508

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

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

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

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

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

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

Army

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

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

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

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

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

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

1 DR. JAMES L. RANEY
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

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

Army

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

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

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

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

Air Force

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

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

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

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

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

Marines

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

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

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

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

CoastGuard

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

Other DoD

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

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

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

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

1 DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civil Govt

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

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

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

Non Govt

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

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

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

1 Capt. J. Jean Belanger
Training Development Division
Canadian Forces Training System
CFTSHQ, CFB Trenton
Astra, Ontario KOK 1B0

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

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

Non Govt

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

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

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

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

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

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

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

Non Govt

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

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

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

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

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

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

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

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

Non Govt

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

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

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

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

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 0854

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

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

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

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