A Comparison of Computer Mastery Models When Pool Characteristics Vary

Robert L. Smith and Charles Lewis

Educational Testing Service, Princeton, New Jersey

Paper presented at the annual meeting of the National Council on Measurement in Education

New Orleans, April 2002

A Comparison of Computer Mastery Testing Models When Pool Characteristics Vary

Introduction

In computer mastery testing the issue of the optimal classification method has periodically been examined. Kingsbury and Weiss (1983) compared the sequential probability ratio test (SPRT) procedure with a sequential adaptive Bayes (SAB) procedure (referred to as Adaptive Mastery Testing (AMT) by Kingsbury and Weiss). The SPRT method used in the Kingsbury and Weiss study delivers a single fixed sequence of items to all test takers using a binomial probability model and a likelihood ratio-based stopping rule, where only the test length varies among test takers. The sequential adaptive Bayes procedure is adaptive in the sense that item selection is individualized for a given test taker. The method is IRT-based with items selected to maximize information at the provisional ability estimate. Because the set of items administered to an individual may vary in the amount of information provided, $1-\alpha$ percent Bayesian confidence bands (using Owen's (1969, 1975) restricted Bayesian updating procedure) are constructed around the posterior mean using the posterior standard deviation for ability after item *i*. If the cut score is not contained in the interval, testing stops and the individual is classified as a master or nonmaster. Otherwise, testing continues up to some maximum test length. If the maximum test length is reached, testing stops and the individual is classified based on the relative relationship of the estimated ability to the cut score. The Kingsbury and Weiss results suggested that the Bayesian procedure was more efficient than the SPRT procedure. However, the methods were not strictly comparable as implemented since a binomial probability model was used with the SPRT procedure and an IRT probability model was used with the SAB.

Spray and Reckase (1996) compared two similar methods for computer mastery testing. In their study, the likelihood ratios used by the SPRT procedure are computed using an IRT probability model rather than the binomial probability model used by Kingsbury and Weiss (1983). The single fixed sequence of items presented to every test taker was ordered with respect to the amount of information at the cut score ability level. Their version of Kingsbury and Weiss' Bayesian procedure used the Bayesian confidence interval stopping rule, but it did not select items adaptively. Instead, the same fixed item presentation sequence chosen for the SPRT procedure was used here as well. This modification will be referred to as a sequential Bayesian (SB) procedure (not adaptive with regard to item selection). With average classification error rates matched, average test lengths could be compared to determine which method was more efficient. Spray and Reckase found the SPRT method required fewer items, on average, than the sequential Bayes method.

Vos and Glas (2000) compared variations of what they referred to as adaptive sequential mastery tests (ASMTs) with sequential Bayesian mastery tests similar to those studied by Lewis and Sheehan (1990) and Smith and Lewis (1995, 1996, 1997, 1998), where item (or testlet) selection was random. In the first of their variations, items were selected that maximized information at a cut score (similar to the SPRT method). A second variation selected items that maximized information at the provisional estimate of ability (similar to AMT). The third variation used Bayesian decision theory to select items and was primarily directed at balancing losses due to misclassification with the cost of testing. Items were selected such that the variance of the difference between mastery and non-mastery losses was minimized. This is a Bayesian decision rule that considers information at both the current ability estimate and the cut score. They found little difference among the methods when the Rasch model was used, but found the third variation of the ASMT method to be more efficient when the 3PL model was used. One limitation of this study is that the item parameters were independently drawn from a standard normal distribution for each simulated test taker, thus, mimicking item selection from an infinite item pool. This situation may mask effects that occur due to the limitations found in finite item pools. For example, selecting items at a test taker's provisional ability estimate may be most efficient when decisions are made quickly based on short tests. However, as test length increases, more items that are increasingly distant from the provisional ability estimate are likely to be administered, thus degrading the benefit of selecting items that are most appropriate for the test taker. Here the finiteness of the item pool interacts with the selection algorithm.

The studies above differ in ways that make direct comparisons difficult. The major issues seem to revolve around the algorithm used for item selection (maximum information at the cut score or maximum information at the ability estimate) and the stopping rule (SPRT or Bayesian confidence intervals). The present study seeks to compare these methods of item selection and stopping rules under various pool configurations and cut scores .

Method

Procedures investigated

Two item selection criteria were examined: 1) selecting items that maximize information at the cut score, and 2) selecting items that provide maximum information at the provisional ability estimate. In addition, two stopping rules were examined, 1) the likelihood ratio and 2) Bayesian posterior probability. The crossing of these factors produced four methods:

1. Sequential Likelihood Ratio Method (SLR) – Likelihood ratio stopping rule, with items selected to have maximum information at the cut score.

2. Sequential Bayes Method (SB) – Bayes posterior probability stopping rule with items selected at the cut score.

3. Sequential Adaptive Likelihood Ratio method (SALR) - Likelihood ratio stopping rule, with items selected to have maximum information at the provisional ability estimate.

4. Sequential Adaptive Bayes (SAB) method - Bayes posterior probability stopping rule with items selected to have maximum information at the provisional ability estimate.

The item selection rules and the stopping rules were separated so that cross combinations could be examined[1].

The Bayesian posterior probability method investigated here is modified slightly from the one proposed by Kingsbury and Weiss (1983). Kingsbury and Weiss used Owen's (1969, 1975) restricted Bayesian updating procedure, where simplifying normal posterior forms were assumed. The present version uses the actual posterior distribution of an individual's ability estimate with a non-informative prior. For the SAB a 99 percent confidence interval around the provisional ability estimate was used. This was operationalized in the following manner. If 99.5 percent of a test taker's posterior probability fell above the cut score, she was passed. Similarly, if 99.5 percent of the test taker's posterior probability fell below the cut score, she was failed.

The likelihood ratio tests were investigated setting the hypothesis error rates equal ($\alpha = \beta$), with the indifference region around the cut score (d) set at ± .20[2]. The values for α (and β) were adjusted to match error rates for SAB, and thus, will change from one condition to another. (See appendix for more details on both stopping rules.)

Pool characteristics

The simulations conducted for this study utilized item parameter estimates from a large pool of items assembled for a certification test using computerized adaptive delivery Four different sub-pools of items were sampled from the large pool. All sub-pools contained 251 items. This size was chosen to be roughly representative of pool sizes for many actual certification tests. The advantage of sampling all sub-pools from a larger pool of items is that it circumvents scaling issues that would come into play if the items were taken from different tests independently scaled. It also allowed for better control of average item discrimination across conditions. (See Table 1 for summary statistics for the different item distributions.)

The item difficulties in the original pool were well distributed across all ability levels between ± 3.0. The a-parameters in this pool were fairly well distributed conditional on item difficulty with the following exception. For easier items, the conditional distributions of the a-parameters tended to have fewer highly discriminating items. This characteristic of the larger pool was reflected in the sub-pools where many items were drawn from the easier portion of the item difficulty distribution. (For example, see Table 2 for the normal, peaked at -.60, and uniform distributions.)

The distributional form of the item difficulties in the sub-pools was as follows (also see Table 3):

[1] Option 3, the SALR, has never been explored to our knowledge and we are not necessarily advocating its use. We include it for completeness.

[2] Other hypothesis error rate and indifference region combinations could have been used to arrive at the same classification error rates. This occurs because the hypothesis error rate and the indifference region can offset each other, e.g., raising the α and β error rates while shrinking the width of the indifference region by a comparable amount will yield the same average misclassification error rate.

1. An approximately normal distribution of item difficulties (b-parameters) centered at an ability of -.55

2. An approximately normal distribution of item difficulties (b-parameters) centered at an ability of .01

3. A truncated distribution of difficulty composed of the 251 most difficult items from the base pool

4. An approximately uniform distribution of item difficulties (b-parameters) between ± 3.0

Simulation Procedures

The test taker ability distribution is taken from an actual testing population for a certification test. The distribution is unimodal, centered approximately at a theta of zero. The distribution of test taker ability was fixed for all methods and sub-pools investigated. For each method and sub-pool three cut scores were investigated: -.60, .00, and 1.3. These corresponded to population pass rates of 72.4, 49.4 and 9.5, respectively. A minimum test length of one item and a maximum test length of 90 items was adopted. The delivery unit was limited to the item. Content constraints were intentionally ignored because they would obscure the results of the study. Similarly, exposure control was not used since it would tend to obscure any differences between the methods. We followed Spray and Reckase (1996) by matching the procedures on average classification error rate, then examining average test length.

Simulations

Data were generated for 10,000 simulated examinees according to the distribution of examinees found in the testing population. Specifically, 10,000 values of θ were drawn from the population distribution. For each θ, 251 item responses were generated using the 3PL model and item parameter estimates form the sub-pool. A different data set was generated for each pool configuration. Each method was applied to the same data set within a pool configuration.

Matching error rates

The matching of error rates was accomplished by first establishing the error rate for the Sequential Adaptive Bayes (SAB) method with α set equal to .01. The other methods were then matched to this error rate by adjusting the α level through an iterative process.

Results

Table 4 presents the overall classification error rates, and average test length by pool distribution, cut score, item selection method and stopping rule. In general, the classification error rates tend to be higher when the cut score is near the center of the ability distribution, where more test takers are located, and lower when the cut score is further out in the tail of the distribution, where fewer test takers are found.

Testing efficiency

Normal item distribution centered at -.55. For the normal item distribution centered at -.55, the likelihood ratio stopping rule had the shortest average test length for all three cut scores. Moreover, item selection method had little effect on mean test length for this stopping rule. For a comparable item selection method, the likelihood ratio stopping rule achieved shorter test lengths than the posterior probability rule by a difference of between 7 and 14 items.

For the posterior probability method, selecting items at the cut score (SB) produced the shorter mean test length when the cut score was below the mean (-.60, 72.4 percent pass rate). However, when the other cut scores (.00 and 1.3) were used, selecting the items at the ability (SAB) produced shorter test lengths by 4 and 13 items, respectively).

Normal item distribution centered at .01. For the normal item distribution centered at .01 the SLR method had the shortest average test length for all three cut scores. The SAB method produced the longest tests at a comparable classification error rate. For a comparable item selection method, the likelihood ratio stopping rule achieved shorter test lengths than the posterior probability rule by a difference of between 4 and 11 items.

Truncated item distribution. For the truncated item distribution the SLR method again had the shortest average test length for all three cut scores. For the extreme cut score (1.3), selecting items at the cut score produced shorter tests, on average, than selecting items at the ability estimate regardless of stopping rule. At this cut there was little difference between the SB, SALR and SAB methods. Controlling for item selection method, the likelihood ratio stopping rule produced shorter average test lengths than the posterior probability method. Differences ranged from less than one item (1.3 cut score, with item selection at ability) to almost 14 items (-.60 cut score, with item selection at the cut score).

Uniform item distribution. The pattern observed for the uniform distribution is similar to the other sub-pools for the lowest cut score (-.60). The likelihood ratio stopping rule obtains shorter average test lengths than the posterior probability stopping rule when method of item selection is controlled for. Also, item selection at the cut score produces shorter average test lengths than when items are selected at the ability.

Some interesting limitations of the methods are observed for the middle cut score (.00). While the SLR method still produces the shortest average test length, the SAB method obtains the second shortest average test length of those that could be evaluated. In addition, for the SALR method it was not possible to reduce the classification error rate to be comparable to the other methods even for average test lengths approaching the maximum test length. (The mean test length listed applies to the minimum attainable error rate which is also given.)

At the extreme cut score, the SAB method produced the shortest average test length. When items were selected at the cut score, neither of the stopping rules could obtain sufficiently low classification error rates to be comparable to the conditions when the items were selected at the ability. In this case, as in the others like this, lowering alpha levels produced tests that had longer average test lengths, but did not improve classification error rates.

Stage of classification

For each combination of pool distribution, cut score and method, Table 5 shows the stage (item) when the first decision was made, the percentage of decisions made through the tenth stage (item 10), and the percentage of decisions made after a test of maximum length (90 items. In general, the posterior probability stopping rule tends to begin classifications a couple of stages earlier than the likelihood ratio stopping rule. The SALR method tends to make many fewer decisions through stage 10 than the other methods. However, by the last stage of testing, when a classification decision is required, both likelihood ratio methods make many fewer decisions than the posterior probability methods. For the uniform sub-pool (where applicable), both likelihood ratio methods tend to make fewer early and fewer late decisions than the posterior probability methods.

Discussion

Two methods of item selection (maximum information at the cut and maximum information at the ability estimate) crossed with two stopping rules (likelihood ratio and posterior probability) for a variable-length mastery test were examined for four different configurations of item sub-pools (normal centered at -.55, normal centered at .01, a truncated sub-pool of the most difficult items, and a uniformly distributed sub-pool. For both of the normal and the truncated sub-pools, the SLR proved to be the most efficient method for cut scores of -.60 and .00, yielding the shortest average test lengths after controlling for classification error rates. At the most extreme cut score (1.3), the SLR method produced the shortest average tests in the normal (.01) and truncated sub-pools, but was found to do no better than the SALR method in the normal (-.55) sub-pool. The posterior probability stopping rule generally tended to be less efficient for these sub-pools.

The pattern observed for the uniform distribution is similar to the other sub-pools for the lowest cut score (-.60). The likelihood ratio stopping rule obtained shorter average test lengths than the posterior probability stopping rule when method of item selection was controlled for, and item selection at the cut score produced shorter average test lengths than when items were selected at the ability estimates. At the middle cut score (.00), the SLR method was the most efficient. The other methods did not differ from each other. At the extreme cut score (1.3) the SAB method produced the shortest average test length. When items were selected at the cut score for this case, neither of the stopping rules could obtain sufficiently low classification error rates to be comparable to the conditions when the items were selected at the ability. This was also observed at the cut score of .00 for the SALR method. In these cases, lowering alpha levels produced tests that had longer average test lengths, but did not improve classification error rates

The different methods showed different patterns of the stage at which test takers were classified. The posterior probability stopping rule tended to begin classifications a couple of stages earlier than the likelihood ratio stopping rule. The SALR method tended to be conservative in the early stages, classifying many fewer test takers through stage 10 than the other methods. However, at the last stage of testing, both likelihood ratio methods make many fewer decisions than the posterior probability methods. For the uniform sub-pool (where applicable), both likelihood ratio methods tended to make fewer early and fewer late decisions than the posterior probability methods.

References

Kingsbury, G.G. & Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D.J. Weiss (Ed.), *New horizons in testing*. New York: Academic Press.

Lewis, C. & Sheehan, K. (1990). Using Bayesian decision theory to design a computer mastery test. *Applied Psychological Measurement*, *14*, 367-386.

Owen, R. J. (1969). *A Bayesian approach to tailored testing*. (Research Bulletin 69-92). Princeton, New Jersey: Educational Testing Service.

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

Smith, R. & Lewis, C. (1995, April). *A Bayesian computerized mastery model with multiple cut scores*. Paper presented at the annual meeting of NCME, San Francisco.

Smith, R. & Lewis, C. (1996, April). *A search procedure to determine sets of decision points when using testlet-based Bayesian sequential testing procedures*. Paper presented at the annual meeting of NCME, New York.

Smith, R. & Lewis, C. (1997, April). *Incorporating decision consistency into Bayesian sequential testing*. Paper presented at the annual meeting of NCME, Chicago.

Smith, R. & Lewis, C. (1998, April). *Expected losses for individuals in Computerized Mastery Testing*. Paper presented at the annual meeting of NCME, San Diego.

Spray, J. A. & Reckase, M. D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*, 405-414.

Vos, H. J. & Glas, C. A. W. (2000). Testlet-based adaptive mastery testing. In W. J. van der Linden and C.A.W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Kluwer Academic Publishers.

Appendix

The Likelihood Ratio Decision Criterion

The likelihood function for θ based on a response pattern

$$\boldsymbol{u}' = (u_1, u_2, \cdots, u_n)$$

for n items may be written as

$$L(\theta \mid \boldsymbol{u}) = \prod_{i=1}^{n} P_i(\theta)^{u_i} [1 - P_i(\theta)]^{1-u_i} ,$$

where $P_i(\theta)$ denotes the item response function for the ith item to be administered.

If our cut score on the θ scale is θ_c, we define two hypothesis values for θ:

$$\theta_0 = \theta_c - d$$

and

$$\theta_1 = \theta_c + d ,$$

with $d > 0$ specified as a minimum difference from the cut score. In other words, the interval from θ_0 to θ_1 is considered to be an indifference region. The likelihood ratio criterion is then defined as

$$LR(\boldsymbol{u}) = L(\theta_1 \mid \boldsymbol{u}) / L(\theta_0 \mid \boldsymbol{u}) .$$

For values of this criterion sufficiently greater than unity, we will pass a candidate, while for values sufficiently close to zero the candidate would be failed. In this study, we treat the two types of classification errors symmetrically. In this case, the critical values for the likelihood ratio test may be written as a function of α_{LR}, a theoretical error rate at the two hypothesis values, θ_0 and θ_1 for a sequentially administered test of potentially unlimited length:

$$LR_0 = \alpha_{LR} / (1 - \alpha_{LR})$$

and

$$LR_1 = (1 - \alpha_{LR}) / \alpha_{LR} .$$

The Likelihood Ratio decision rule after administering n items is

1. Pass if $LR(\mathbf{u}) > LR_1$;
2. Fail if $LR(\mathbf{u}) < LR_0$;
3. Otherwise, continue testing.

If the maximum test length has been reached, the final decision rule is given by

1. Pass if $L(\theta_1 \mid \mathbf{u}) \geq L(\theta_0 \mid \mathbf{u})$;
2. Fail if $L(\theta_0 \mid \mathbf{u}) > L(\theta_1 \mid \mathbf{u})$.

(In practice, for reasons of convenience and numerical precision, all tests are carried out using the natural logarithm of the likelihood ratio.)

The Posterior Probability Decision Criterion

We begin by identifying a set of quadrature points on the θ scale:

θ_q^*, for $q = 1, \cdots, Q$.

We introduce a prior, or population probability distribution for θ on these points, denoted by $p\left(\theta_q^*\right)$. After administering n items, and observing a response vector u, we compute a posterior distribution for θ using Bayes' Theorem:

$$p(\theta \mid \boldsymbol{u}) = \frac{L(\theta \mid \boldsymbol{u})p(\theta)}{\sum_{q=1}^{Q} L\left(\theta_q^* \mid \boldsymbol{u}\right)p\left(\theta_q^*\right)}.$$

Using this distribution, we can find the posterior probability that θ will lie below (or above) the cut score θ_c:

$$\Pr(\theta < \theta_c \mid \boldsymbol{u}) = \sum_{q:\theta_q^* < \theta_c} p\left(\theta_q^* \mid \boldsymbol{u}\right)$$

and

$$\Pr(\theta \geq \theta_c \mid \boldsymbol{u}) = 1 - \Pr(\theta < \theta_c \mid \boldsymbol{u}).$$

To describe a symmetric decision rule based on this posterior probability, we first introduce a theoretical error rate, α_B. The Posterior Probability decision rule after administering n items may then be formulated as follows:

1. Pass if $\Pr(\theta < \theta_c \mid \mathbf{u}) < \alpha_B / 2$;

2. Fail if $\Pr(\theta \geq \theta_c \mid \mathbf{u}) < \alpha_B / 2$;

3. Otherwise, continue testing.

(This rule is comparable to the more traditional formulation: evaluating the location of θ_c relative to a $(1-\alpha_B)\times 100\%$ Bayesian confidence interval. It has the advantage over that approach of allowing the easy introduction of unequal error rates for the two types of classification errors, as can also be done for the Likelihood Ratio procedure.)

If the maximum test length has been reached, the final decision rule is given by

1. Pass if $\Pr(\theta \geq \theta_c \mid \mathbf{u}) \geq$ ½;

2. Fail if $\Pr(\theta < \theta_c \mid \mathbf{u}) >$ ½.

Although a population prior distribution for θ was available for our study, and was used in the simulations to draw samples of θ values, we used a uniform prior ($p(\theta_q^*) = 1/Q$ for all q) to compute the posterior distributions for this decision rule. This choice is consistent with our earlier work on Bayesian mastery testing, and may be justified on grounds of fairness to candidates.

Table 1

Summary Statistics for Item Parameters in the Various Pools

	a		b		c	
Distribution	Mn	SD	Mn	SD	Mn	SD
Normal (-.55)	.70	.19	-.55	.69	.22	.06
Normal (.01)	.72	.20	.01	.71	.21	.06
Truncated	.73	.19	.70	.73	.20	.06
Uniform	.69	.20	-.11	1.39	.20	.06

Table 2

Distribution of a-Parameters by Pool Configuration

a-parameter	Normal (-.55)	Normal (.01)	Truncated	Uniform
.35	17	12	8	25
.45	26	31	30	29
.55	29	25	29	28
.65	58	48	45	56
.75	41	39	43	32
.85	40	42	43	40
.95	25	33	32	21
1.05	15	21	21	20

Table 3

Distribution of b-Parameters by Pool Configuration

b-parameter	Normal (-.55)	Normal (.01)	Truncated	Uniform
-2.9	0	0	0	4
-2.7	0	0	0	0
-2.5	0	0	0	8
-2.3	1	0	0	8
-2.1	2	0	0	5
-1.9	4	1	0	8
-1.7	7	2	0	11
-1.5	11	3	0	11
-1.3	14	4	0	11
-1.1	20	8	0	11
-.9	26	14	0	11
-.7	26	20	0	11
-.5	29	23	0	11
-.3	29	27	26	11
-.1	21	21	21	11
.1	22	28	28	11
.3	15	25	25	11
.5	10	23	24	11
.7	7	20	23	11
.9	4	14	24	11
1.1	2	8	20	11
1.3	1	4	15	11
1.5	0	3	14	11
1.7	0	2	9	9
1.9	0	1	6	6
2.1	0	0	9	9
2.3	0	0	4	4
2.5	0	0	1	1
2.7	0	0	0	0
2.9	0	0	2	2

Table 4

Classification Error Rates and Mean Test Lengths by Item Selection Method and Stopping Rule

Sub-pool Item Distribution	Cut Score (theta)	Item Selection (Max Info)	Stopping Rule	Name	Classification Error (%)	Mean Test Length	Alpha
Normal Centered at -.55	-.60	Cut	LR	SLR	5.01	24.57	.1180
	-.60	Cut	Post. Prob.	SB	5.01	34.88	.0118
	-.60	Ability	LR	SALR	5.01	25.70	.1300
	-.60	Ability	Post. Prob.	SAB	5.01	38.01	.0100
	.00	Cut	LR	SLR	5.75	31.12	.0950
	.00	Cut	Post. Prob.	SB	5.75	44.81	.0040
	.00	Ability	LR	SALR	5.76	31.71	.0850
	.00	Ability	Post. Prob.	SAB	5.75	40.97	.0100
	1.3	Cut	LR	SLR	3.30	19.54	.1230
	1.3	Cut	Post. Prob.	SB	3.30	39.62	.0001
	1.3	Ability	LR	SALR	3.30	19.33	.1410
	1.3	Ability	Post. Prob.	SAB	3.30	26.27	.0100
Normal Centered at .01	-.60	Cut	LR	SLR	5.23	25.08	.1210
	-.60	Cut	Post. Prob.	SB	5.23	36.16	.0102
	-.60	Ability	LR	SALR	5.23	31.00	.0100
	-.60	Ability	Post. Prob.	SAB	5.24	39.27	.0100
	.00	Cut	LR	SLR	5.99	27.99	.0950
	.00	Cut	Post. Prob.	SB	5.98	36.48	.0130
	.00	Ability	LR	SALR	5.97	31.59	.0850
	.00	Ability	Post. Prob.	SAB	5.99	41.02	.0100
	1.3	Cut	LR	SLR	2.81	14.42	.1420
	1.3	Cut	Post. Prob.	SB	2.81	19.30	.0180
	1.3	Ability	LR	SALR	2.80	19.70	.1000
	1.3	Ability	Post. Prob.	SAB	2.82	23.92	.0100

Table 4 (continued)

Classification Error Rates and Mean Test Lengths by Item Selection Method and Stopping Rule

Sub-pool Item Distribution	Cut Score (theta)	Item Selection (Max Info)	Stopping Rule	Name	Classification Error (%)	Mean Test Length	Alpha
Truncated Dist.	-.60	Cut	LR	SLR	6.55	26.87	.1500
	-.60	Cut	Post. Prob.	SB	6.56	40.57	.0104
	-.60	Ability	LR	SALR	6.55	32.95	.1390
	-.60	Ability	Post. Prob.	SAB	6.56	43.54	.0100
	.00	Cut	LR	SLR	6.21	27.33	.1075
	.00	Cut	Post. Prob.	SB	6.21	36.78	.0137
	.00	Ability	LR	SALR	6.21	30.82	.1070
	.00	Ability	Post. Prob.	SAB	6.21	41.99	.0100
	1.3	Cut	LR	SLR	2.54	18.04	.0900
	1.3	Cut	Post. Prob.	SB	2.54	22.33	.0090
	1.3	Ability	LR	SALR	2.54	22.47	.0650
	1.3	Ability	Post. Prob.	SAB	2.54	22.86	.0100
Uniform	-.60	Cut	LR	SLR	5.92	28.84	.1195
	-.60	Cut	Post. Prob.	SB	5.93	39.16	.0100
	-.60	Ability	LR	SALR	5.92	30.05	.1380
	-.60	Ability	Post. Prob.	SAB	5.92	41.71	.0100
	.00	Cut	LR	SLR	6.32	36.44	.0700
	.00	Cut	Post. Prob.	SB	6.32	47.29	.0045
	.00	Ability	LR	SALR	6.35*	43.27*	.0500
	.00	Ability	Post. Prob.	SAB	6.32	44.11	.0100
	1.3	Cut	LR	SLR	2.91*	22.16*	.0700
	1.3	Cut	Post. Prob.	SB	2.90*	33.08*	.0010
	1.3	Ability	LR	SALR	2.85	26.88	.0500
	1.3	Ability	Post. Prob.	SAB	2.86	24.36	.0100

* Results are misleading due to inability to match on classification error rates. Tests of infinite length would not have resulted in matched classification error rates.

Table 5

Percentage of Decisions at Early and Late Stages of Testing

Sub-pool Item Distribution	Cut Score (theta)	Item Selection (Max Info)	Stopping Rule	Name	Stage of First Decision	Decisions through Stage 10 (percent)	Decisions at Last Stage (90) (percent)
Normal	-.60	Cut	LR	SLR	5	32.76	6.45
Centered	-.60	Cut	Post. Prob.	SB	4	39.06	23.20
at -.55	-.60	Ability	LR	SALR	5	7.71	5.66
	-.60	Ability	Post. Prob.	SAB	4	33.37	25.62
	.00	Cut	LR	SLR	8	12.75	10.74
	.00	Cut	Post. Prob.	SB	6	23.40	37.07
	.00	Ability	LR	SALR	6	8.87	10.36
	.00	Ability	Post. Prob.	SAB	3	31.10	29.04
	1.3	Cut	LR	SLR	5	60.91	7.28
	1.3	Cut	Post. Prob.	SB	3	62.59	29.44
	1.3	Ability	LR	SALR	4	55.37	6.27
	1.3	Ability	Post. Prob.	SAB	2	57.19	17.88
Normal	-.60	Cut	LR	SLR	6	33.63	7.45
Centered	-.60	Cut	Post. Prob.	SB	4	39.37	25.29
at .00	-.60	Ability	LR	SALR	7	1.58	9.61
	-.60	Ability	Post. Prob.	SAB	4	28.39	27.04
	.00	Cut	LR	SLR	7	19.96	8.25
	.00	Cut	Post. Prob.	SB	4	36.91	25.28
	.00	Ability	LR	SALR	7	9.73	9.71
	.00	Ability	Post. Prob.	SAB	3	28.79	28.53
	1.3	Cut	LR	SLR	5	64.97	2.86
	1.3	Cut	Post. Prob.	SB	3	68.24	11.86
	1.3	Ability	LR	SALR	5	38.25	4.98
	1.3	Ability	Post. Prob.	SAB	2	56.37	14.39

Table 5 (continued)
Percentage of Decisions at Early and Late Stages of Testing

Sub-pool Item Distribution	Cut Score (theta)	Item Selection (Max Info)	Stopping Rule	Name	Stage of First Decision	Decisions through Stage 10 (percent)	Decisions at Last Stage (90) (percent)
Truncated Dist.	-.60	Cut	LR	SLR	7	32.61	9.33
	-.60	Cut	Post. Prob.	SB	6	34.80	29.84
	-.60	Ability	LR	SALR	10	.70	10.59
	-.60	Ability	Post. Prob.	SAB	4	22.26	30.87
	.00	Cut	LR	SLR	7	28.06	8.55
	.00	Cut	Post. Prob.	SB	5	36.85	26.22
	.00	Ability	LR	SALR	9	1.89	8.98
	.00	Ability	Post. Prob.	SAB	3	26.09	29.28
	1.3	Cut	LR	SLR	6	50.86	4.40
	1.3	Cut	Post. Prob.	SB	4	61.04	13.20
	1.3	Ability	LR	SALR	7	21.50	5.88
	1.3	Ability	Post. Prob.	SAB	4	59.52	13.28
Uniform	-.60	Cut	LR	SLR	5	6.07	9.98
	-.60	Cut	Post. Prob.	SB	4	38.25	28.11
	-.60	Ability	LR	SALR	4	4.38	8.60
	-.60	Ability	Post. Prob.	SAB	3	26.75	29.75
	.00	Cut	LR	SLR	9	2.07	15.41
	.00	Cut	Post. Prob.	SB	7	19.43	36.34
	.00	Ability	LR	SALR	*	*	*
	.00	Ability	Post. Prob.	SAB	3	25.63	32.92
	1.3	Cut	LR	SLR	*	*	*
	1.3	Cut	Post. Prob.	SB	*	*	*
	1.3	Ability	LR	SALR	8	7.82	8.80
	1.3	Ability	Post. Prob.	SAB	2	55.64	15.22

* Results are misleading due to inability to match on classification error rates.