AAS TECHNICAL NOTE

Smart Test Technology as a Computer Adaptive Item Selection Algorithm

Damon U. Bryant, PhD

Summary: The purpose of this research was to demonstrate the utility of Adaptive Assessment Services' Smart Test Technology ® proprietary algorithm (STT) as an item selection algorithm for computer adaptive testing. There were three independent variables in this study: item selection method, size of the item pool, and length of the adaptive test. The three item selection methods were random item selection, STT, and maximum information, $I(\theta)_{max}$. The size of the item pool was 200, 300, or 400 items. The length of the adaptive test was 10, 15, 20, 25, or 30 questions. Each condition in the 3 x 3 x 5 design was replicated with 200 examinees for a total of 9000 simulations. The dependent variables were bias of ability (θ), mean squared error of θ, item pool utility rate, and maximum item exposure rate. Results suggest that STT (1) is unbiased compared to random and $I(\theta)_{max}$ item selection methods, (2) is more accurate than random item selection in estimating θ, (3) is comparable in accuracy to $I(\theta)_{max}$ when 20 or more items are administered in an adaptive test, (4) is two to four times more efficient in utilizing the entire item pool as compared to $I(\theta)_{max}$ item selection, (5) is equally efficient in utilizing the entire item pool as compared to random item selection, (6) yields maximum item exposure rates that are significantly lower than $I(\theta)_{max}$ exposure rates without using any item exposure control mechanisms, and (7) yields item exposure rates that are orthogonal to item parameters of exposed questions. The evidence in this research strongly suggests that STT is the best item selection algorithm for computer adaptive testing.

Purpose and Objectives

Purpose: To explore the utility of Smart Test Technology ® (STT) as a computer adaptive item selection algorithm.

Objective 1: To show that STT used as a computer adaptive item selection algorithm is more accurate in estimating ability as compared to random item selection.

Objective 2: To show that STT used as a computer adaptive item selection algorithm is comparable in accuracy to computer adaptive item selection using maximum information.

Objective 3: To show that STT yields item exposure rates that are orthogonal to item parameter estimates.

Objective 4: To show that STT is more efficient than maximum information in the utilization of the item pool.

Objective 5: To show that STT yields lower maximum item exposure rates than maximum information.

Introduction

Lord (1970) describes an individual being measured efficiently as follows: “An examinee is measured most effectively when the test items are neither too difficult nor too easy” (p. 139). Despite popular use of maximum item information, $I(\theta)_{max}$, in the design and administration of computer adaptive tests, from a theoretical standpoint, it is not the concept that most accurately represents the effectiveness of measuring psychological attributes of examinees according to the aforementioned statement. Moreover, from a practical point of view, the use of $I(\theta)_{max}$ as an item selection method has major shortcomings. Specifically, the exposure rate of questions selected using $I(\theta)_{max}$ is highly correlated with the discrimination parameter of questions within the item pool for advanced models, e.g., unidimensional 2- or 3-Parameter Logistic Model (PLM). Thus, items with high discrimination parameters are overexposed, and items with low discrimination parameters are underexposed. As a consequence, the utilization rate of the item pool, i.e., the number of exposed items divided by the total number of items in the pool, is unacceptably low. This is a threat to test security and poor utilization of the item pool. This has spawned new research in the area of item exposure control.

Although dozens of item exposure control methods have been proposed (See Georgiadou, Triantafillou, & Economides, 2007) as an adjustment to $I(\theta)_{max}$, these procedures also have several limitations in terms of efficiency in application. First, most methods require the test developer to conduct computer simulations of item pools prior to operational rollout to ensure reasonable exposure rates; the computations are time-intensive and cumbersome. Second, the psychometrician must then select arbitrary thresholds for determining overexposure. Third, the correlation between item exposure rate and the discrimination parameter in using $I(\theta)_{max}$ is still present but may be attenuated to a certain extent. Fourth, despite the control of overexposed questions, there still exists in some cases the problem of poor utilization of the item pool and underexposed questions. These theoretical and applied problems revolve around the criteria of maximizing item information.

Based upon the aforementioned challenges with $I(\theta)_{max}$ and subsequent item exposure control algorithms, some criteria for a more effective item selection method for computer adaptive testing become evident: First, an acceptable item selection algorithm should more accurately match the concept of effective measurement according to Lord. Second, the item selection algorithm should not only yield item exposure rates that are relatively low and evenly distributed across the item pool, but also yield exposure rates that are orthogonal to item parameters, specifically item discrimination (a_i). Third, the item selection algorithm should yield a very high item pool utilization rate, thus maximizing all available operational resources and increasing test security. Fourth, the item selection method should be more effective than random item selection in estimating θ and comparable to item selection based upon $I(\theta)_{max}$. It should be noted that maximizing information cannot be improved upon in terms of accuracy as this method seeks to minimize error in estimating θ. However, we argue that the proposed method should be relatively accurate when compared to $I(\theta)_{max}$. To the degree that a method is proposed that meets these criteria, then an improvement over $I(\theta)_{max}$ without item exposure control mechanisms has been achieved. The purpose of this paper is to advance an item selection method that achieves all of the aforementioned goals.

This study investigates the extent to which the proprietary Smart Test Technology ® (STT) is efficient in selecting questions and estimating θ in a test as compared to $I(\theta)_{max}$ and random item selection. STT does not use any item exposure control mechanisms. $I(\theta)_{max}$ and random item selection are compared to STT because they do not implement any item exposure control mechanisms.

Method

This study was conducted via computer simulation. Question probabilities were generated according to the 3-Parameter Logistic Model (3-PLM). Item difficulties (b_i) were created via a normal distribution with a mean of 0 and standard deviation of 1. Item discrimination (a_i) values for the 400 questions were generated via a normal distribution with a mean of 1.2 and a standard deviation of .25. The pseudo-guessing parameter (c_i) was generated according to a uniform distribution with a minimum value of .00 and maximum value of .27. Ability parameters were generated by a normal distribution with a mean of 0 and standard deviation of 1. The number of examinees simulated within each condition described below is 200.

The independent variables are as follows: (1) test length (10, 15, 20, 25, and 30 questions administered), (2) item selection method [random, STT, and $I(\theta)_{max}$], and (3) item pool size (200, 300, 400 questions). This is a 5x3x3 fixed effect design. The same simulated examinees were used throughout all conditions in the study for comparability of results. The dependent variables include the following: (1) bias, (2) mean squared error of θ, (3) utility rate of item pool, and (4) maximum item exposure rate. Bias is operationalized as the difference between the estimated ability parameter (θ_{EST}) and true ability parameter (θ). Mean squared error of θ (MSE) is the mean squared difference between estimated ability parameters and true ability parameters. Utility rate of the item pool is operationally defined as the number of items within an item pool that have been exposed to examinees divided by the total number of items in the pool. A value of 100% indicates that every item in a pool has been administered in operation at least once to examinees. A value of 50% indicates that half of the items in a pool have been administered at least once to examinees. Higher values indicate greater utility. The last dependent variable is maximum item exposure rate, which is the percentage value of the item with the highest exposure rate within each one of the 45 conditions in this simulation. High values indicate an undesirable exposure rate; low values indicate a low exposure rate. The standard of evidence in this study is $\alpha = .01$.

Results

There was no significant difference among the item selection methods in bias of θ, $F(44, 8955) = 1.04$, p = ns. In other words, all item selection methods have relatively small amounts of bias. The averages of bias across all conditions were -.015, .005, and -.011 for random, STT, and $I(\theta)_{max}$ item selection methods, respectively. As a consequence, this result is no longer discussed.

The first objective of this study was to show that STT used as a computer adaptive item selection algorithm is more accurate in estimating θ as compared to random item selection. The results of the analysis of variance (ANOVA) indicated that there was a significant difference among one or more of the items selection methods, $F(2, 8955) = 141.32$, $p \leq .001$. The effect size of this difference is $\eta^2 = .05$, accounting for 5% of the variance in MSE. Over the conditions investigated in this study, the random item selection method had the largest amount of error in scores ($M = .155$, $SD = .249$), followed by STT ($M = .109$, $SD = .202$) and $I(\theta)_{max}$ ($M = .072$, $SD = .119$). Post-hoc tests showed that STT item selection had significantly lower error in estimates of θ than random item selection. $I(\theta)_{max}$ had a significantly lower error in estimating θ than STT ($p \leq .001$). However, these differences were considered to be conditional effects in the presence of a statistically significant interaction of item length and item selection method, $F(8, 8955) = 4.41$, $p \leq .001$. This interaction was plotted and graphically displayed in Figure 1. As can be seen in Figure 1, when the number of items administered in a computer adaptive test increased, the mean squared error decreased across item selection methods. However, the decrease in error was particularly salient for the STT item selection method. As more items were administered, the rate of decrease in error was greater than random item selection and $I(\theta)_{max}$.

In an effort to evaluate the notion that STT and $I(\theta)_{max}$ are comparable in accuracy, post-hoc tests were performed. Bonferroni post-hoc tests showed that when 10 items were administered, there was a significant difference in error between random item selection ($M = .248$) and STT ($M = .203$), $p \leq .005$. In other words, when 10 items were administered, STT was more accurate in estimating theta. At 10 item administered, $I(\theta)_{max}$ item selection method had a significantly lower mean squared error ($M = .126$) than both random and STT item selection methods (all *p*s $\leq .001$) . However, as the number of items administered increased, the difference in error between STT and $I(\theta)_{max}$ was no longer statistically significant. This was evident when 20, (.092 vs. .060, p = ns), 25 (.069 vs. .045, p = ns) and 30 (.056 vs. .043, p = ns) items were administered. In other words, there was no difference in the accuracy between STT and $I(\theta)_{max}$ when 20 or more items were administered in a computer adaptive test.

Another objective of this study was to show that STT yields item exposure rates that are orthogonal to item parameter estimates. To evaluate this proposition, the correlations among item parameters and exposure rate for each of the three item selection methods were computed and displayed in Table 2. As can be seen in Table 2, the item parameters of discrimination (a_i), difficulty (b_i), and pseudo-guessing (c_i) were not significantly correlated with the exposure rate of items administered using random item selection or STT (all *p*s = ns). However, when the item selection method was $I(\theta)_{max}$, the item parameter values were significantly correlated with the exposure rate. As expected, the item exposure rate of $I(\theta)_{max}$ was positively correlated with item discrimination ($r = .57$, $p \leq .001$) with approximately 30% shared variance. Moreover, the item exposure rate was negatively correlated with the pseudo-guessing parameter ($r = -.29$, $p \leq .001$). In other words, when using $I(\theta)_{max}$, as an item selection method in computer adaptive testing, items were often selected if they had high discrimination values and low pseudo-guessing values. However, when using STT or random item selection, there was no relationship between item exposure rates and item discrimination values of exposed items; this was also true with difficulty and pseudo-guessing. In other words, STT item selection method administers items in such as way as to make exposure rates orthogonal to item parameters of discrimination, difficulty, and guessing.

Another objective of this study was to demonstrate how STT yields higher utilization rates of the item pool as compared to $I(\theta)_{max}$. This analysis of variance was conducted by using the utility rate values in Tables 5-7. The independent variables in this analysis were item pool size and item selection method. The items administered variable was collapsed so that there were at least 5 cases within each condition of the 3 x 3 ANOVA in predicting utility rate. The overall model was statistically significant, $F(8,36) = 108.75$, $p \leq .001$, accounting for approximately 96% of the variance in item pool utility rate. As expected, item selection method was statistically significant, $F(2,36) = 427.16$, $p \leq .001$, accounting for approximately 94% of the variance in item pool utility rate. Post-hoc tests indicated that random item selection (M = 99.92) and STT item selection (M = 98.54) did not differ in regard to utilizing the item pool as both methods have utilization rates over 98.5%. However, $I(\theta)_{max}$ item selection was far less effective in utilizing the entire item pool, only using 35.2% of the items in the pool. See Figure 1. This evidence suggests that STT is far more superior than $I(\theta)_{max}$ in utilizing all of the items in an item pool.

The final objective of this study was to show that STT yields lower maximum item exposure rates than $I(\theta)_{max}$. See Table 4. The ANOVA model predicting maximum exposure rate from the item selection method and item pool size was statistically significant, $F(8,36) = 351.66$, $p \leq .001$, accounting for approximately 99% of the variance in maximum item exposure rate. The item selection method was statistically significant in predicting maximum item exposure rate, $F(2,36) = 1394.51$, $p \leq .001$. This variable accounted for approximately 98% of the variance. See Figure 3. For the $I(\theta)_{max}$ selection method, the average maximum exposure rate was 86.9% across all conditions. For the STT selection method, the average maximum exposure rate across all conditions was 23.8%; this difference in the average maximum exposure rates was statistically significant, ($p \leq .001$). The evidence of this analysis strongly suggests that STT has a maximum item exposure rate that is dramatically lower than the maximum item exposure rate of $I(\theta)_{max}$.

Discussion

The purpose of this research was to demonstrate the efficiency of Adaptive Assessment Services Smart Test Technology ® (STT) in the selection of items, the estimation of ability, and the utilization of the item pool. The primary objective of this study was show that STT is just as accurate as $I(\theta)_{max}$ and more accurate than random item selection in the estimation of θ. A second objective was to show that

STT used within a computer adaptive algorithm is more efficient in its use of the item pool and has lower maximum item exposure rates as compared to $I(\theta)_{max}$.

In regard to accuracy, the results of this study show that STT is just as accurate as $I(\theta)_{max}$ in the estimation of θ when 20 or more items are administered. There was no reliable difference in mean squared error between the two adaptive item selection algorithms. STT and $I(\theta)_{max}$ were both more accurate than random item selection, which was expected given the theoretical benefits of adaptive testing. When bias was used as a criterion for evaluating STT, $I(\theta)_{max}$, and random item selection, there was virtually no bias in the estimation of θ. This evidence suggests that STT as an adaptive algorithm with no item exposure control is equivalent in accuracy to $I(\theta)_{max}$ in estimating θ when 20 or more items are administered.

In terms of utilization of the item pool for adaptive testing, STT had a significantly higher utilization rate than $I(\theta)_{max}$. The utilization rate was over 98%, which means that virtually all of the items were used. This is comparable to using a random item selection algorithm, which was shown to have a utilization rate of 99%. In contrast, the utilization rate of $I(\theta)_{max}$ was less than 50% (i.e., 35%), which is unacceptable considering the costs in developing an operational item pool. Because of the limited number of conditions in this study, caution should be used to interpret these results as replications are needed to support these findings in more general settings. However, within the context of this study which covers most testing conditions, evidence suggests that STT used as an adaptive algorithm is two times more efficient in utilizing the full item pool as compared to $I(\theta)_{max}$.

In terms of individual item exposure rates, STT had an average maximum individual item exposure rate of 25% across the conditions in this study. $I(\theta)_{max}$ had an average maximum individual item exposure rate over 85%, which means that some items were exposed to almost every person who took an adaptive test. The maximum item exposure rate for random item selection was approximately 15%. The individual item exposure rate of STT was significantly lower than the individual item exposure rate of $I(\theta)_{max}$. This evidence suggests that STT as an item selection algorithm substantially reduces the maximum item exposure rate of individual items as compared to $I(\theta)_{max}$ with no item exposure control mechanism. Future research should determine the extent to which these results generalize to a variety of other conditions.

The evidence in this study paints a very dark picture for the use of $I(\theta)_{max}$ as an item selection algorithm in a computer adaptive test when there is no item exposure controls. Although $I(\theta)_{max}$ is just as accurate as STT in estimating θ at 20 or more items, it has the lowest item pool utilization rate, highest mean item exposure rate, and highest correlation among item parameters and exposure rates. In an operational testing program, the implications are clear: Unless efficient item exposure controls are built into the $I(\theta)_{max}$ adaptive algorithm, the time and money used to develop an operational item pool are wasted due to half the items not even being exposed in an administration.

In contrast, STT is just as accurate as $I(\theta)_{max}$ in estimating theta when 20 or more items are administered, has a significantly higher item pool utilization rate, a significantly lower mean item exposure rate, and produces no reliable relationships among item parameters and exposure rates of those same items. It is very interesting to note that in terms of the utilization rate of the item pool, mean item exposure rate, and variance of item exposure, STT as a computer adaptive algorithm performs very similar to a random item selection algorithm in that the utilization rate of the item pool is near 100%, the mean item exposure rate approaches the expectation of the exposure rate given the number of items administered and the size of the item pool, and the variance of the exposure rate is at a minimum. Although this finding is what one would want ideally in a computer adaptive context, this may not be the case in all circumstances. When the distribution of the difficulty parameters does not match the ability distribution of the examinee population, for example, item exposure rates may be correlated with the difficulty parameter but not discrimination; this would be the case for any adaptive item selection algorithm. If the goal of test design is to estimate the ability distribution of a sample of examinees from a population, then the distribution of difficulty estimates should be designed to match the distribution of the population from which the examinee sample comes. In most testing contexts, it is implied that questions of appropriate difficulty are constructed to match the ability distribution of the examinee population.

In summary, the purpose of this study was to investigate STT as a computer adaptive testing algorithm. It was compared to random and $I(\theta)_{max}$ item selection methods with no item exposure control mechanisms. Results suggest that STT (1) is unbiased compared to random and $I(\theta)_{max}$ item selection methods, (2) is more accurate than random item selection in estimating θ, (3) is comparable in accuracy to $I(\theta)_{max}$ when 20 or more items are administered in an adaptive test, (4) is two to four times more efficient in utilizing the entire item pool as compared to $I(\theta)_{max}$ item selection, (5) is equally efficient in utilizing the entire item pool as compared to random item selection, (6) yields maximum item exposure rates that are significantly lower than $I(\theta)_{max}$ exposure rates without using any item exposure control mechanisms, and (7) yields item exposure rates that are orthogonal to item parameters of exposed questions. Future research can determine the extent to which the findings generalize to the multidimensional case. There is no reason to suggest that results may differ in a multidimensional context.

References

Georgiadou, E., Triantafillou, E., Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment, 5*(8). Retrieved 5/4/2010 from http://www.jtla.org.

Lord, F. M. (1970). Some test theory for tailored testing. In W. H. Holtzman (Ed.), *Computer-assisted instruction, testing, and guidance* (pp. 139–183). New York: Harper & Row.

Table 1. Analysis of Variance Predicting Mean Squared Error of θ

Factor	*MS*	*df*	*F*		η^2
Item Pool (IP)	.00	2	.09		.00
Test Length (TL)	4.91	4	133.93	**	.05
Item Selection Method (SM)	5.18	2	141.32	**	.03
IP x TL	.03	8	.84		.00
IP x SM	.04	4	1.15		.00
TL x SM	.16	8	4.41	**	.01
IP x TL x SM	.04	16	1.02		.00

Overall Model: $F(44, 8955) = 20.03$ **, $R^2 = .09$

* $p < .01$, ** $p < .001$

Table 2. Correlation of Item Exposure Rate and IRT Parameters

Item Selection Method		ER	*a*	*b*	*c*
Random	ER	-			
N=4495	*a*	.00	-		
	b	.00	.02	-	
	c	.03	**.05**	**-.05**	-
Smart Test Technology	ER	-			
N=4426	*a*	.00	-		
	b	-.03	.02	-	
	c	.01	**.05**	**-.05**	-
Maximum Information	ER	-			
N=1498	*a*	**.57**	-		
	b	-.04	.04	-	
	c	**-.29**	**.13**	.00	-

Bold = $p \leq .01$, ER = Exposure Rate

Item was included in analysis if it was selected by one of the three methods above.

Table 3. Analysis of Variance Predicting Item Pool Utility Rate

Factor	*MS*	*df*	*F*		η^2
Item Pool (IP)	.02	2	3.16		.01
Item Selection Method (SM)	2.05	2	427.16	**	.94
IP x SM	.01	4	2.34		.01

Overall Model: $F(8, 36) = 108.75$ **, $R^2 = .96$

* $p < .01$, ** $p < .001$

Table 4. Analysis of Variance Predicting Maximum Item Exposure Rate

Factor	*MS*	*df*	*F*		η^2
Item Pool (IP)	.02	2	9.90	**	.01
Item Selection Method (SM)	2.40	2	1394.51	**	.98
IP x SM	.00	4	1.12		.00

Overall Model: $F(8, 36) = 351.66$ **, $R^2 = .99$
* $p < .01$, ** $p < .001$

Table 5. Exposure Statistics for the Simulated Conditions (Item Pool Size = 200)

Item Pool	Test Length	Item Selection Method	*MSE* (θ)	Item Pool Utility Rate (%)	Item Exposure Rate (*M*)	Item Exposure Rate (*SD*)	Item Exposure Max.	Item Exposure Min.
200	10	Random	.2270	100.0	.0500	.0151	.0900	.0150
		STT	.2057	97.0	.0515	.0381	.2000	.0050
		$I(\theta)_{max}$	.1416	26.5	.1887	.2154	.8500	.0050
	15	Random	.1856	100.0	.0750	.0191	.1300	.0300
		STT	.1457	99.5	.0754	.0483	.2750	.0050
		$I(\theta)_{max}$	.0944	36.5	.2055	.2129	.8800	.0050
	20	Random	.1353	100.0	.1000	.0214	.1850	.0550
		STT	.0898	100.0	.1000	.0596	.3150	.0100
		$I(\theta)_{max}$	.0628	45.5	.2198	.1985	.9050	.0050
	25	Random	.1170	100.0	.1250	.0209	.1800	.0700
		STT	.0564	100.0	.1250	.0663	.3200	.0150
		$I(\theta)_{max}$	.0465	53.5	.2336	.1971	.8700	.005
	30	Random	.0787	100.0	.1500	.0252	.2250	.0800
		STT	.0555	100.0	.1500	.0724	.3900	.0150
		$I(\theta)_{max}$	.0420	60.5	.2479	.1918	.9250	.0050

Table 6. Exposure Statistics for the Simulated Conditions (Item Pool Size = 300)

Item Pool	Test Length	Item Selection Method	*MSE* (θ)	Item Pool Utility Rate (%)	Item Exposure Rate (*M*)	Item Exposure Rate (*SD*)	Item Exposure Max.	Item Exposure Min.
300	10	Random	.2347	100.0	.0333	.0137	.0700	.0050
		STT	.1975	94.0	.0353	.0259	.1500	.0050
		$I(\theta)_{max}$	.1232	18.7	.1786	.2118	.8550	.0050
	15	Random	.1758	100.0	.0500	.0158	.1000	.0200
		STT	.1321	99.7	.0502	.0344	.1950	.0050
		$I(\theta)_{max}$	.0787	26.7	.1875	.2053	.8650	.0050
	20	Random	.1577	100.0	.0667	.0183	.1200	.0200
		STT	.0922	99.3	.0671	.0430	.2100	.0050
		$I(\theta)_{max}$	.0667	34.3	.1941	.1961	.8850	.0050
	25	Random	.1136	100.0	.0833	.0195	.1400	.0350
		STT	.0884	100.0	.0833	.0473	.2300	.0050
		$I(\theta)_{max}$	.0464	40.7	.2049	.1862	.8300	.0050
	30	Random	.0943	100.0	.1000	.0217	.1600	.0450
		STT	.0546	100.0	.1000	.0548	.2950	.0100
		$I(\theta)_{max}$	.0422	47.7	.2098	.1830	.8800	.0050

Table 7. Exposure Statistics for the Simulated Conditions (Item Pool Size = 400)

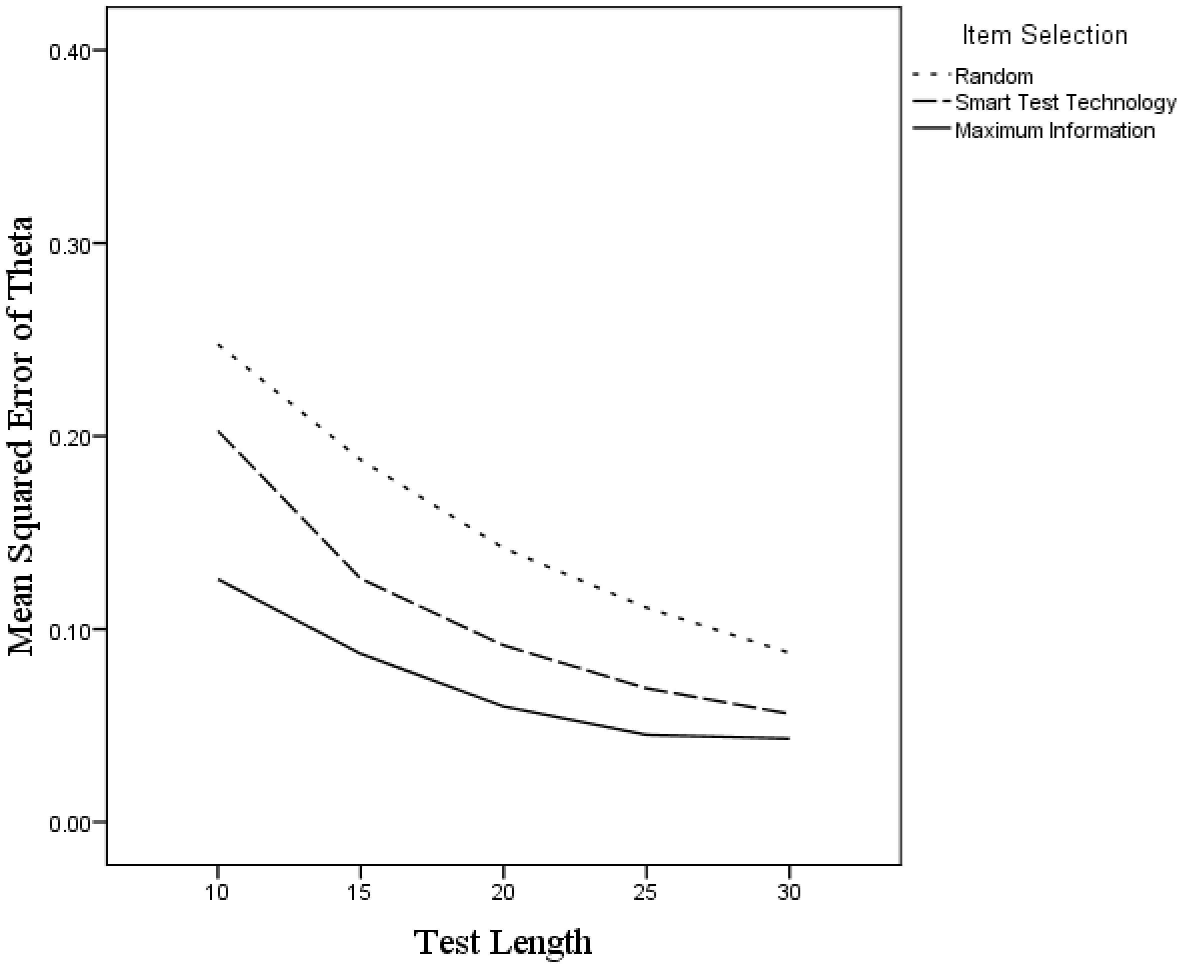
Item Pool	Test Length	Item Selection Method	*MSE* (θ)	Item Pool Utility Rate (%)	Item Exposure Rate (*M*)	Item Exposure Rate (*SD*)	Item Exposure Max.	Item Exposure Min.
400	10	Random	.2812	98.8	.0253	.0110	.0650	.0050
		STT	.2050	92.5	.0270	.0217	.1500	.0050
		$I(\theta)_{max}$	.1127	14.2	.1754	.2010	.8350	.0050
	15	Random	.2016	100.0	.0375	.0137	.0900	.0100
		STT	.1004	97.2	.0386	.0276	.1600	.0050
		$I(\theta)_{max}$	.0883	20.8	.1807	.1995	.8500	.0050
	20	Random	.1328	100.0	.0500	.0162	.1100	.0050
		STT	.0926	99.2	.0504	.0350	.2150	.0050
		$I(\theta)_{max}$	.0500	28.5	.1754	.1898	.8650	.0050
	25	Random	.1025	100.0	.0625	.0174	.1100	.0150
		STT	.0630	99.2	.0630	.0411	.2350	.0050
		$I(\theta)_{max}$	.0425	34.0	.1838	.1809	.8800	.0050
	30	Random	.0899	100.0	.0750	.0182	.1250	.0300
		STT	.0581	100.0	.0750	.0447	.2350	.0050
		$I(\theta)_{max}$	.0456	39.8	.1887	.1750	.8550	.0050

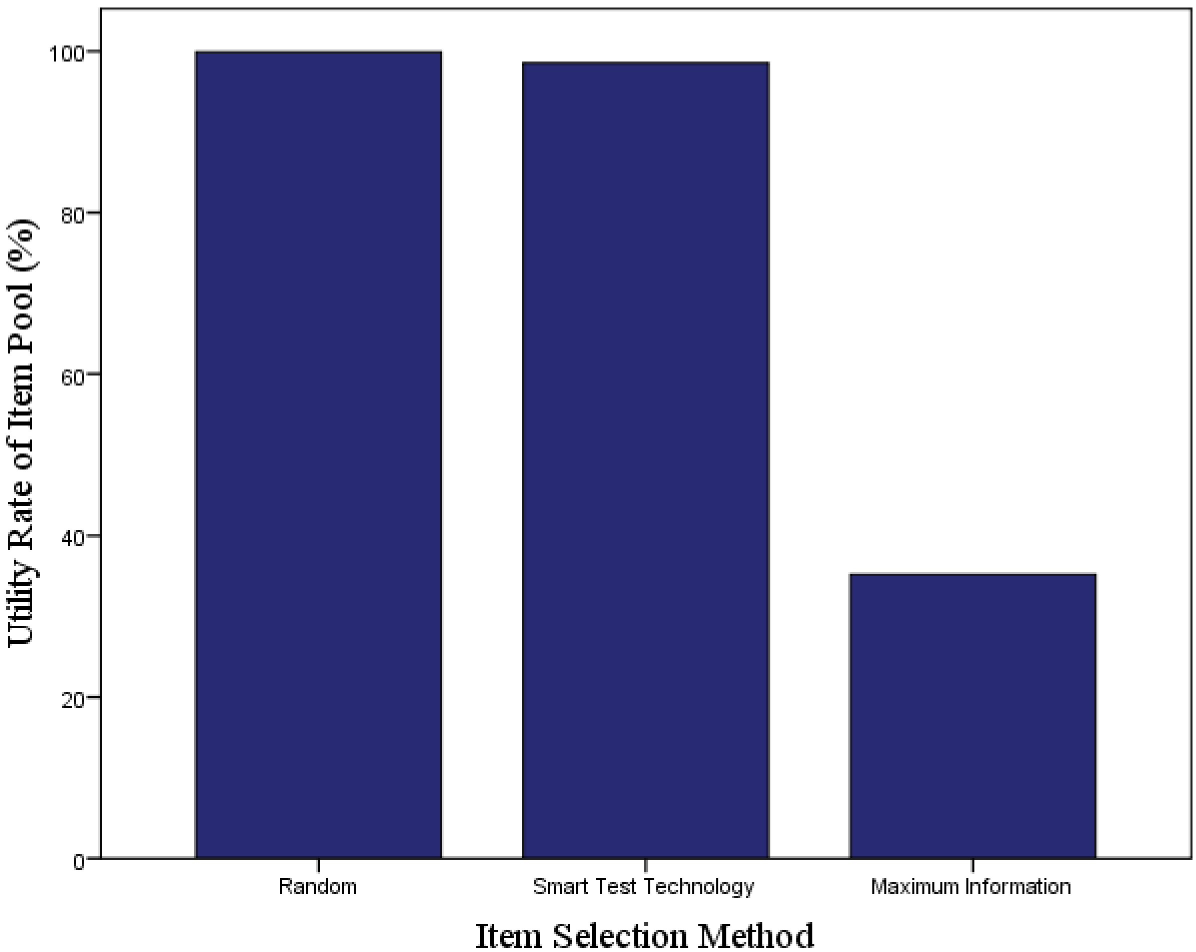
Figure Caption

Figure 1. Interaction of Item Selection Method and Test Length in Predicting Mean Squared Error of θ.

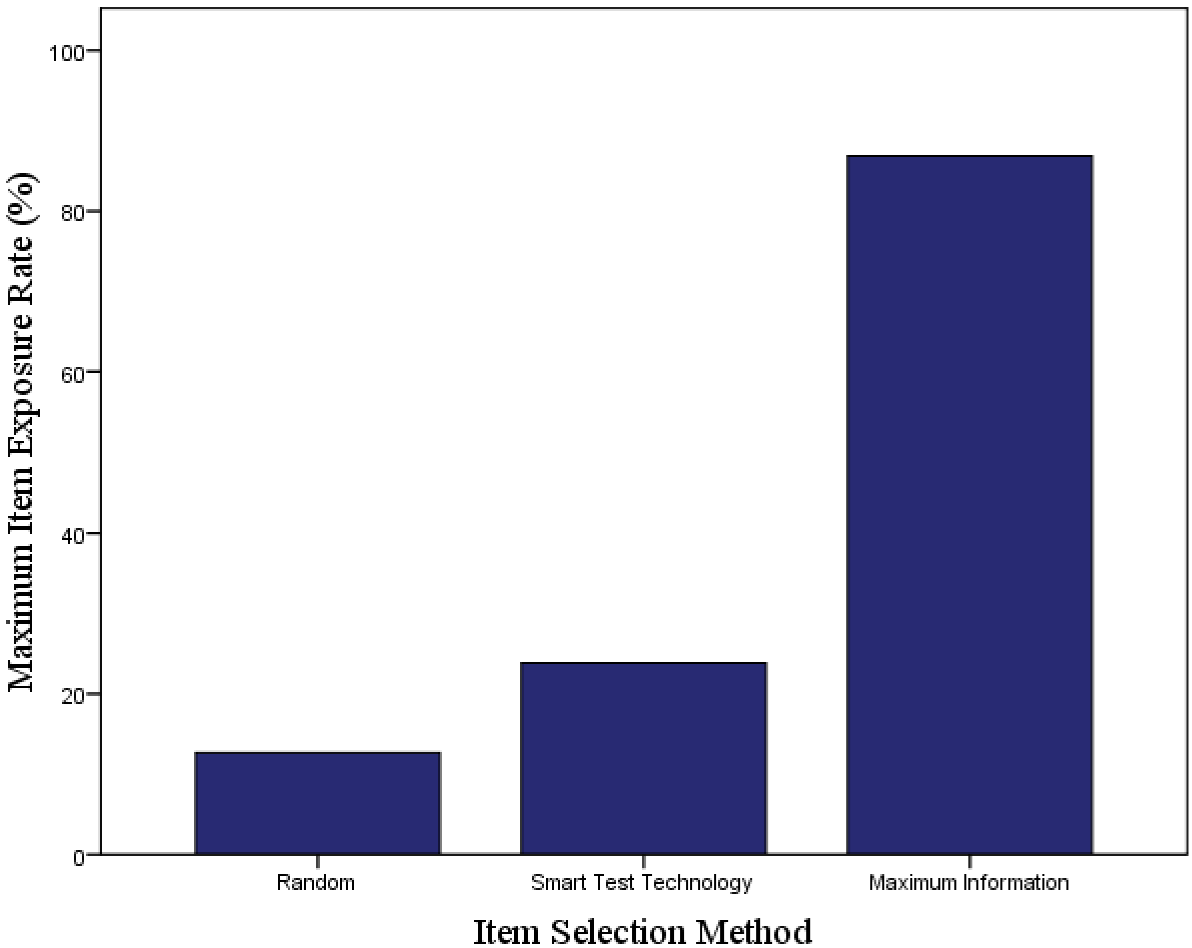
Figure 2. Mean Item Pool Utility Rate as a Function of Item Selection Method.

Figure 3. Mean Maximum Item Exposure Rate as a Function of Item Selection Method


Item Selection
Random
Smart Test Technology
Maximum Information
Mean Squared Error of Theta
0.40
0.30
0.20
0.10
0.00
10
15
20
25
30
Test Length



Utility Rate of Item Pool (%)
100
80
60
40
20
0
Random
Smart Test Technology
Maximum Information
Item Selection Method


100
80
60
40
20
0
Maximum Item Exposure Rate (%)
Random
Smart Test Technology
Maximum Information
Item Selection Method