Computerized Classification Testing under Practical Constraints with a Polytomous Model

C. Allen Lau
Harcourt Brace Educational Measurement

Tianyou Wang
ACT

Paper presented at the 1999 AERA Annual Meeting in Montreal, Canada, April 1999.

Please make correspondence to:
C. Allen Lau, Project Director
Harcourt Brace Educational Measurement
555 Academic Court
San Antonio, Texas 78204
E-mail: allen_lau@hbtpc.com

Computerized Classification Testing under Practical Constraints
with a Polytomous Model

Sequential probability ratio testing (SPRT) procedure was found promising for making mastery decisions in computerized classification testing (CCT) with tests containing dichotomous items (Spray & Reckase, 1996). Lau & Wang (1998) found that SPRT could be applied using the generalized partial credit model. The purposes of this study are to extend the SPRT procedure with the polytomous model under some practical constraints in CCT, such as methods to control item exposure rate and to study the effects of other variables, including item information algorithms, test difficulties, item pool sizes and widths of indifference region in SPRT.

Mastery testing is used to classify the test takers into one of two categories: mastery (pass) or non-mastery (fail). Certification or licensure testing is a good example of it. When such tests are administered and scored in computer format, it is referred to as computerized classification testing (CCT) (Spray, Abdel-fattah, Huang, & Lau, 1997). To implement an IRT-based CCT procedure, a cut-point on the ability scale (θ_c) must be established first. Two types of classification errors are considered: if the examinee is classified as a master but in fact his/her ability level (θ) is below θ_c, a false positive error (type I error) occurs; if the examinee is classified as a nonmaster but in fact his/her θ is at or above θ_c, a false negative error (type II error) occurs. The relative importance of these two types of error is situation dependent.

In CCT, SPRT procedure was found promising for mastery classification (Spray & Reckase, 1996, Lau, 1996, Lau & Wang, 1998). Wald (1947) first proposed the SPRT procedure to test two simple hypotheses: H_0: $P=P_0$ versus H_1: $P=P_1$ with a binomial model. Reckase (1983) modified the procedure and applied it to CCT with IRT models. With SPRT, items are selected to maximize information at the cut-point. Decisions are based on the ratio of the likelihood of the response data conditioned at two alternative points (θ_0 and θ_1) around the cut-point (θ_c) on the θ scale. The interval between these θ_0 and θ_1 is called the indifference region. The width of the indifference region can be set arbitrarily. The decision about the examinee's status (pass or fail) is made based on the consideration of two simple hypotheses:

$$H_0: \theta_j = \theta_0 \quad \text{versus} \quad H_1: \theta_j = \theta_1$$

where θ_j is an unknown parameter, and θ_0 and θ_1 are the lower and upper limits of the indifference region.

Conditioned at these two points, we have $\pi(\theta_1)$ and $\pi(\theta_0)$, where $\pi(\theta_j)$ = Prob ($X = x \mid \theta = \theta_j$), x = 0, 1 are the response data. The product, $\pi_1(\theta_j)\, \pi_2(\theta_j) \ldots \pi_n(\theta_j)$ is called the likelihood function of the response vector. A ratio of these two functions, $L(x) = \pi(\theta_1)/\pi(\theta_0)$, is called a likelihood ratio and

$$L = L(x_1, x_2, \ldots, x_n \mid \theta_0, \theta_1) = \frac{\pi_1(\theta_1)\pi_2(\theta_1)\ldots\pi_n(\theta_1)}{\pi_1(\theta_0)\pi_2(\theta_0)\ldots\pi_n(\theta_0)} \quad .$$

The likelihood ratio is compared to the boundaries, A and B,

where $A = (1-\beta) / \alpha$, and $B = \beta / (1-\alpha)$, and α and β are the error probabilities defined as follows:

Prob(choosing $H_1 \mid H_0$ is true) = α (false positive), and Prob(choosing $H_0 \mid H_1$ is true) = β (false negative).

The likelihood ratio is compared to A and B to make decisions. If $L \geq A$, the H_1 is accepted and the examinee is classified as pass. If $L \leq B$, then H_0 is accepted, and the examinee is classified as fail. If $B < L < A$, then the test continues.

Few if any research investigates how to apply polytomous models in computerized adaptive test (CAT) because of the difficulty of item scoring of the extended response items. Bennett, Steffen, Singley, Morley, & Jacquemin (1997) however, successfully adopted computer scoring of open-ended format items in CAT, which implies the feasibility of polytomous scoring in CCT in the future. Lau & Wang (1998) found that SPRT procedure could be adapted with polytomous items in CCT. Specifically, they found: (a) SPRT procedure with polytomous item pool achieved better classification accuracy than that with dichotomous item; and (b) comparing to partly and totally random item selection, best classification accuracy and efficiency was gained when items were picked based on item information at the cutting point.

This study applied SPRT for polytomous items under Muraki's (1992) generalized partial credit model (GPCM). Under GPCM, the probability of getting a response category h on item i is

$$P_{ih}(\theta) = \frac{\exp\left[\sum_{v=1}^{h} Z_{iv}(\theta)\right]}{\sum_{c=1}^{m_i} \exp\left[\sum_{v=1}^{c} Z_{iv}(\theta)\right]}$$

where $h = 1, 2, \ldots, m$.

within an item, $\sum P_{ih}(\theta) = 1$ *and* $Z_{ih}(\theta) = Da_i(\theta - b_{ih}) = Da_i(\theta - b_i + d_h)$

where

D is a scaling constant that puts the θ ability scale in the same metric as the normal ogive model (D=1.7),

a_i is a slope parameter,

b_{ih} is an item-category parameter,

b_i is an item-location parameter, and

d_h is a category parameter.

The computation of the likelihood ratio for polytomous items is quite similar to the dichotomous SPRT except that the polytomous item response model instead of the dichotomous response model is used to compute the conditional probability of the response data.

Eggen (1998) compared Fisher (F) with Kullback-Leibler (K-L) information (Cover & Thomas, 1991) for item selection in the context of SPRT using a dichotomous item pool. He concluded that the performance of the testing algorithms with K-L were sometimes better and never worse than that of F information-based item selection. In theory, K-L information is more suitable for statistical testing because it is defined as the log of the ratio of two likelihood functions. It seems to be particularly appropriate for SPRT. This study extent this comparison with polytomous item pool.

For dichotomous items, the K-L item information index is defined as:

$$K_i(\theta_1 \| \theta_0) = p_i(\theta_1)\log\frac{p_i(\theta_1)}{p_i(\theta_0)} + q_i(\theta_1)\log\frac{q_i(\theta_1)}{q_i(\theta_0)}$$

For polytomous items, the K-L item information index is:

$$K_i(\theta_1 \| \theta_0) = \sum_{i=0}^{n} p_i(\theta_1)\log\frac{p_i(\theta_1)}{p_i(\theta_0)}$$

where i = 0, 1, 2, ..., n.

Item exposure rate control is important for high stake tests like certificate testing. In CCT, items are usually selected according to the maximum information at the cutting points with SPRT procedure because it guarantees best classification accuracy and efficiency. However, this practice may cause the problem of item over exposure. This study adopted two popular item exposure control methods, Sympson and Hetter method (SH) (Sympson and Hetter, 1985), and Randomesque method (RD) (Kingsbury & Zara, 1989).

As it was mentioned above, the width of the indifference region in SPRT can be set arbitrarily. In theory, the width of the region can affect the number of items used to make mastery decision. Further, the width has an effect to K-L information algorithm, which could impact the testing result. This study tried to investigate how the width of the indifference region affects the results.

Test difficulty and item pool size are practical also constraints in testing and can have an effect on testing results. They were included as independent variables in this study.

Methods

Theoretical method was used to analyze the decision criterion for the polytomous SPRT procedure and to derive possible alternative criterion. Monte Carlo simulation technique was adopted to verify the decision criterion. Several independent variables were manipulated which included:

1. Item information algorithm:
    (1) Fisher.
    (2) Kullback-Leibler.
2. Item exposure control methods:
    (1) Sympson and Hetter method. (Maximum exposure rate was set at 0.25)
    (2) Randomesque method. (For every 3 most informative items unconsidered in the pool, randomly select one item.)
    (3) No control. (The items were only ranked at the cutting theta according to the item information.)
3. Location of theta cut point (test difficulty):
    (1) $\theta_c = -0.8$.
    (2) $\theta_c = 0.8$.

4. Item pool size

(1) 266 items.

(2) 90 items (These 90 items were randomly drawn from the first pool.)

5. Width of Indifference region in SPRT:

(1) $|\theta_0 - \theta_1| = 0.5$ (i.e., $\theta_0 = \delta - 0.25$, $\theta_1 = \delta + 0.25$).

(2) $|\theta_0 - \theta_1| = 1.0$ (i.e., $\theta_0 = \delta - 0.5$, $\theta_1 = \delta + 0.5$).

where δ is the passing criterion.

This was a 2x2x3x3x2 crossed factorial design and these were 48 combinations of conditions totally. Test length constraint (that is, the examinees must respond to a minimum number of items and not exceed a maximum number of items) was set minimum = 3, maximum = 30.

The evaluative criteria include: (1) classification accuracy in terms of false positive and false negative error rates, (2) test efficient (number of items used to make mastery decision), (3) item exposure rate, and (4) item utilization rate. (1 – percentage of not-used items in the item pool)

<u>Data</u>

Item parameters from the 1996 NAEP Science assessment were used to build the item pool. Combining three grades (4th, 8th and 12th) together, the assessment consists 266 polytomous item parameters for the study. These item parameters across three grades were calibrated on the same scale. The average item difficulty of the pool was 1.043. Item response data were generated for 10,000 simulated examinees from a normal distribution (0, 1) on computer.

<u>Steps for Simulation</u>

1. Items were calibrated and ranked at the cutting theta (-0.8 or 0.8) with either Fisher or Kullback-Leibler information algorithm with the two item pools (266 and 90).
2. Item selection was based on Sympson and Hetter, Randomesque method, or no exposure control.
3. 10,000 simulated examinees were administrated and SPRT procedure with different indifference regions was adopted to make mastery decision.
4. Test length, error and item exposure rate were recorded or computed.

Results

The results are listed in Tables 1 to 5. Tables 1 and 2 show the results of item exposure control with Sympson-Hetter and Randomesque methods. Table 3 describes the result of no exposure control condition. Tables 4 and 5 summarize the average error rates, average test lengths, and average item exposure rates and item utilization rates of each manipulated variable across all conditions.

Item Information Algorithm

Two information indexes used for item selection were Fisher and Kullback-Leibler. Amazingly, under different conditions, the results from either information algorithm were very similar. Within each condition and across all conditions, the average type I errors, type II errors, total errors and test lengths were almost identical. (See Tables 1-4.) The average type I, type II, total error, and test length were 0.028, 0.032, 0.061, and 9.326 for Fisher and 0.028, 0.033, 0.061, and 9.333 for Kullback-Leibler. Not only that, the item exposure rates and patterns for both Item information algorithm were again almost identical. (See Table 5.)

As the results of F information were very similar to those of K-L in terms of accuracy, efficiency, and item exposure rate, K-L could be an alternative for item information algorithm in computerized classification testing.

Item Exposure Control Methods

Two popular item exposure methods, Sympson and Hetter, and Randomesque were applied in this study. Across all conditions, SH and RD methods gained similar results in accuracy and efficiency. (See Table 4.) The average type I, type II, total error, and test length were 0.029, 0.034, 0.063, and 10.254 for SH and 0.030, 0.035, 0.065, and 10.014 for RD. Compared to the no exposure control condition, both methods only sacrificed a little accuracy and efficiency.

Generally, both methods offered good control over item exposure rate. In both cases, no items were exposed more than 0.5. For SH method, about 1% of the items exposed over 0.3. For the RD method, about 8% of the items exposed over 0.3. So in terms of strict item exposure control, SH seemed better.

In terms of item utilization rate, on the other hand, RD was better than SH. About 67% of items were used with RD method but only 44% items were used with SH methods. (See Table 5.)

Location of Cutting Theta (Test Difficulty)

In this study, test difficulty influenced the test accuracy and efficiency. Within each condition and across all conditions, as the cutting level increased, the total error and item utilization rate decreased. The average type I, type II, total error rate, and test length were 0.027, 0.042, 0.069, and 11.292 for the cutting theta = -0.8 and 0.029, 0.023, 0.052, and 7.629 for the cutting theta = 0.8. The average number of item used for theta = -0.8 was 48% more than that of theta = 0.8.

These results were reasonable because the average item difficulties of the full (266 items) and partial size (90 items) pool were 1.043 and 0.94 respectively. In theory, these items can distinguish the above average examinees better.

Item pool size

Item pool size was found affecting the classification accuracy and test efficiency. Two item pool sizes, 266 item in the first pool and 90 items in the second. The 90 items in the second pool were randomly drawn from the first item pool with similar grade proportion (27%, 37%, and 36% from grades 4, 8, and 12 respectively.)

Within each condition and across all conditions, the larger item pool consistently had better accuracy and efficiency. (See Tables 1-4.) For the smaller pool, about 47% more items were needed to make the mastery decision and about 33% less classification accuracy compared with the larger pool. The explanation was possibly that more good items (informative items at the cutting theta) could be selected and used from the larger item pool and that improved the testing quality.

Width of Indifference Region in SPRT

With the SPRT procedure, the width of indifference region can be varied. It is kind of arbitrary to set up the width. Two width adopted in this study were: $|\theta_0 - \theta_1| = 0.5$ or 1.0.

The width of the indifference region was found affecting item consumption and testing accuracy. The wider the region, the less items were used to make the mastery decision. When the width was set at 0.5, about 84% more items were needed. (See Table 4.)

Generally, in this study, the error rates were smaller when the width was set at 0.5. The type I, type II, and total error were 0.027, 0.030, and 0.058 with the width equal to 0.5 compared to 0.029, 0.035, 0.064 with the width equal to 1.0.

Conclusion

Polytomous items were again found working well with SPRT procedure in CCT in this study. Several variables were manipulated to investigate the impact on the accuracy, efficiency, item exposure and item utilization.

With all these evaluation criteria, Fisher information was found very similar to those of Kullback-Leibler. So K-L could be another option for item information algorithm in computerized classification testing.

The full size pool gained better classification accuracy and significantly reduced the number of item used compared with the smaller pool in this study. It is believed that more informative items could be utilized in the larger pool. So it is in fact that the item quality improves the testing quality.

This study explored item exposure control rates in the context of CCT with polytomous model. Only two popular methods, Sympson-Hetter and Randomesque were adopted. These two methods were found to produce similar results in classification accuracy and testing efficiency but produce different results in item exposure rate and utilization rate. SH was better in strict item exposure control while RD was better in item utilization. It is situation-dependent to decide which criteria, item exposure control or item utilization is more important. The test users should make this decision. There are other item exposure control methods like McBride and Martin method (McBride & Martin, 1983), Progression method (Revuelta, 1995), and Stocking & Lewis conditional multinomial method (Stocking & Lewis, 1995). Different methods for exposure control with polytomous items should be investigated in the future.

It was found that the width of the indifference region had an impact in SPRT on accuracy and efficiency. In this study, when the width was double, item consumption reduced 46% with sacrificing about 0.6% classification accuracy. There seems to be a trade-off between accuracy and efficiency by changing the width. The test users can adjust the width to fulfil the need. More different widths could be set and investigated in future study.

References

Bennett, R. E., Steffen, M., Singley, M. K., Morley, M., & Jacquemin, D. (1997). Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests. *Journal of Educational Measurement, 34*, 162-176.

Eggen, T. J. H. M. (1998). *Item selection in adaptive testing with the sequential probability ratio test*. Measurement and Research Department Report, 98-1. Arnhem: Cito.

Ercikan, K., Burket, G., Julian, M., Link, V., Schwarz, R., & Weber, M. (1996). *Calibration and scoring of tests with multiple-choice and constructed response item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Kalohn, J. C., & Spray, J. A. (1998*). Effect of item selection on item exposure rates within a computerized classification test*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

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: latent trait test theory and computerized adaptive testing*. (pp. 257-283) New York: Academic Press.

Kingsbury, G. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education, 4,* 241-261.

Lau, C. A. (1996*). Robustness of a unidimensional computerized mastery testing procedure with multidimensional testing data.* Unpublished doctoral dissertation, University of Iowa, 1996.

Lau, C. A., & Wang, T. (1998)*. Comparing and combining dichotomous and polytomous items with SPRT procedure in computerized classification testing*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Luecht, R. M. (1998). *A framework for exploring and controlling risks associated with test item exposure over time*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Reckase, M. D. (1983). A procedure for decision making using tailored testing, In D. J. Weiss (Ed.), *New horizons in testing; latent trait test theory and computerized adaptive testing* (pp. 237-255). New York: Academic Press.

Spray, J. A., Abdel-fattah, A. A., Huang, C. & Lau, C. A. (1997). *Unidimensional approximations for a computerized test when the item pool and latent space are multidimensional*. (ACT Research Report Series 97-5). Iowa City, IA: American College Testing.

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.

Spray, J., Reckase, M. D. (1987). *The effect of item parameter estimation error on decisions made using the sequential probability ratio test* (ACT Research Report Series 87-1). Iowa City, IA: American College Testing.

Stocking, M. L & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement, 17*, 277-292.

Stocking, M. L. (1993). Controlling item exposure rates in a realistic adaptive testing paradigm. (ETS Research Report Series). Princeton, New Jersey: Educational Testing Service.

Sympson, J. B. & Hetter, R. D. (1985). Controlling item exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association*, (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.

Wald, A. (1947). *Sequential Analysis*. New York: Dover Publications, Inc.

Table 1. Sympson-Hetter Exposure Control: Errors Rates, Test Length, Pass, and Fail Rates

Cutting Theta	Indifference Region	Pool Size	Inform Algorithm	Type I Error	Type II Error	Total Error	Test Length	Pass Rate	Fail Rate
-0.8	0.5	266	Fisher	0.023	0.032	0.056	12.739	0.778	0.222
0.8	0.5	266	Fisher	0.025	0.018	0.043	8.971	0.211	0.789
-0.8	0.5	90	Fisher	0.035	0.052	0.087	18.465	0.773	0.227
0.8	0.5	90	Fisher	0.038	0.025	0.063	13.699	0.223	0.777
-0.8	0.5	266	K-L	0.022	0.033	0.054	12.804	0.780	0.220
0.8	0.5	266	K-L	0.023	0.019	0.042	8.863	0.208	0.792
-0.8	0.5	90	K-L	0.036	0.054	0.090	18.523	0.762	0.238
0.8	0.5	90	K-L	0.035	0.028	0.063	13.578	0.223	0.777
-0.8	1.0	266	Fisher	0.024	0.037	0.062	6.818	0.772	0.228
0.8	1.0	266	Fisher	0.031	0.027	0.058	4.759	0.220	0.780
-0.8	1.0	90	Fisher	0.028	0.055	0.083	10.404	0.766	0.234
0.8	1.0	90	Fisher	0.034	0.026	0.060	6.439	0.224	0.776
-0.8	1.0	266	K-L	0.023	0.039	0.063	6.693	0.774	0.226
0.8	1.0	266	K-L	0.030	0.024	0.054	4.639	0.223	0.778
-0.8	1.0	90	K-L	0.025	0.047	0.072	10.354	0.768	0.232
0.8	1.0	90	K-L	0.036	0.027	0.063	6.322	0.212	0.788

Note: K-L is the Kullback-Leibler information.

Table 2. Randomesque Exposure Control: Errors Rates, Test Length, Pass, and Fail Rates

Cutting Theta	Indifference Region	Pool Size	Inform Algorithm	Type I Error	Type II Error	Total Error	Test Length	Pass Rate	Fail Rate
-0.8	0.5	266	Fisher	0.025	0.033	0.058	12.492	0.778	0.222
0.8	0.5	266	Fisher	0.024	0.021	0.044	8.734	0.212	0.788
-0.8	0.5	90	Fisher	0.034	0.048	0.082	17.787	0.777	0.223
0.8	0.5	90	Fisher	0.037	0.027	0.064	12.769	0.210	0.790
-0.8	0.5	266	K-L	0.024	0.033	0.057	12.498	0.778	0.222
0.8	0.5	266	K-L	0.021	0.019	0.041	8.709	0.210	0.790
-0.8	0.5	90	K-L	0.033	0.050	0.083	17.835	0.761	0.239
0.8	0.5	90	K-L	0.031	0.026	0.058	12.873	0.214	0.786
-0.8	1.0	266	Fisher	0.026	0.041	0.067	6.786	0.766	0.234
0.8	1.0	266	Fisher	0.033	0.024	0.056	4.742	0.216	0.784
-0.8	1.0	90	Fisher	0.031	0.055	0.086	10.298	0.767	0.233
0.8	1.0	90	Fisher	0.038	0.028	0.066	6.365	0.227	0.773
-0.8	1.0	266	K-L	0.024	0.043	0.067	6.692	0.768	0.232
0.8	1.0	266	K-L	0.029	0.023	0.051	4.852	0.213	0.787
-0.8	1.0	90	K-L	0.031	0.059	0.091	10.419	0.760	0.240
0.8	1.0	90	K-L	0.039	0.029	0.068	6.378	0.221	0.779

Table 3. No Exposure Control: Errors Rates, Test Length, Pass, and Fail Rates

Cutting Theta	Indifference Region	Pool Size	Inform Algorithm	Type I Error	Type II Error	Total Error	Test Length	Pass Rate	Fail Rate
-0.8	0.5	266	Fisher	0.022	0.028	0.050	10.803	0.779	0.221
0.8	0.5	266	Fisher	0.020	0.016	0.037	6.194	0.215	0.785
-0.8	0.5	90	Fisher	0.028	0.036	0.064	13.856	0.785	0.215
0.8	0.5	90	Fisher	0.024	0.022	0.047	8.777	0.213	0.787
-0.8	0.5	266	K-L	0.023	0.028	0.051	10.576	0.776	0.224
0.8	0.5	266	K-L	0.023	0.017	0.040	6.539	0.218	0.782
-0.8	0.5	90	K-L	0.027	0.036	0.063	13.730	0.776	0.224
0.8	0.5	90	K-L	0.025	0.021	0.046	8.780	0.216	0.784
-0.8	1.0	266	Fisher	0.025	0.036	0.061	5.552	0.777	0.223
0.8	1.0	266	Fisher	0.023	0.019	0.042	3.977	0.214	0.786
-0.8	1.0	90	Fisher	0.027	0.042	0.069	7.893	0.773	0.227
0.8	1.0	90	Fisher	0.027	0.024	0.051	4.503	0.211	0.789
-0.8	1.0	266	K-L	0.024	0.039	0.063	5.702	0.772	0.228
0.8	1.0	266	K-L	0.026	0.021	0.048	4.015	0.217	0.783
-0.8	1.0	90	K-L	0.026	0.045	0.071	8.000	0.770	0.230
0.8	1.0	90	K-L	0.029	0.026	0.055	4.622	0.211	0.789

Table 4. Average Error Rates and Test Length of The Independent Variables

Independent Variable	Type I Error	Type II Error	Total Error	Test Length
Item Information Algorithm				
Fisher	0.028	0.032	0.061	9.326
K-L	0.028	0.033	0.061	9.333
Exposure Control Method				
SH	0.029	0.034	0.063	10.254
RD	0.030	0.035	0.065	10.014
No Control	0.025	0.029	0.054	7.720
Cutting Theta				
$\theta_c = -.8$	0.027	0.042	0.069	11.292
$\theta_c = .8$	0.029	0.023	0.052	7.629
Pool Size				
266	0.025	0.028	0.053	7.715
90	0.032	0.037	0.069	11.366
Indifference Region Width				
0.5	0.027	0.030	0.058	12.108
1.0	0.029	0.035	0.064	6.573

Note: K-L is the Kullback-Leibler information. SH is Sympson and Hetter item exposure control method. RD is Randomesque item exposure control method.

Table 5. Average Item Exposure Rates of The Independent Variables

Independent Variable	r=0	0<r<.1	.1≤r<.2	.2≤r<.3	.3≤r<.4	.4≤r<.5	r≥.5
Item Information Algorithm							
Fisher	0.558	0.183	0.089	0.120	0.033	0.005	0.013
K-L	0.557	0.181	0.092	0.120	0.032	0.005	0.012
Exposure Control Method							
SH	0.564	0.101	0.041	0.284	0.005	0.005	0.000
RD	0.331	0.371	0.178	0.043	0.078	0.000	0.000
No Control	0.777	0.074	0.053	0.034	0.014	0.010	0.038
Cutting Theta							
$\theta_c = -.8$	0.544	0.132	0.114	0.148	0.038	0.009	0.016
$\theta_c = .8$	0.571	0.233	0.067	0.092	0.027	0.001	0.009
Pool Size							
266	0.786	0.121	0.032	0.041	0.014	0.001	0.005
90	0.328	0.243	0.150	0.199	0.051	0.009	0.020
Indifference Region Width							
0.5	0.532	0.111	0.121	0.171	0.041	0.008	0.016
1.0	0.583	0.253	0.060	0.069	0.024	0.002	0.009