# Comparing and Combining Dichotomous and Polytomous Items with SPRT Procedure in Computerized Classification Testing

C. Allen Lau
Harcourt Brace Educational Measurement

Tianyou Wang
ACT

Paper presented at the 1998 AERA Annual Meeting in San Diego, California, April 1998.

# Comparing and Combining Dichotomous and Polytomous Items with SPRT Procedure in Computerized Classification Testing

Certification or licensure testing is a kind of mastery testing used to classify the test takers into one of two categories: qualified (pass) or unqualified (fail). When the certification or licensure testing is administered and scored in computer format, it is referred to as computerized classification testing (CCT) (Spray, Abdel-fattah, Huang, & Lau, 1997). The main objective of CCT is to make accurate mastery decisions with lowest possible cost, including the testing time. To implement an item response theory (IRT)-based CCT procedure, a cut-point on the ability scale ($\theta_c$) 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 ($\theta$) is below $\theta_c$, a false positive error (type I error) occurs; if the examinee is classified as a nonmaster but in fact his/her $\theta$ is at or above $\theta_c$, a false negative error (type II error) occurs. The relative importance of these two types of error is situation dependent.

Two main procedures have been proposed to make mastery decisions: (1) sequential Bayes, or SB (Kingsbury & Weiss, 1983), or (2) sequential probability ratio testing, or SPRT (Reckase, 1983; Wald, 1947). Both SB and SPRT utilize IRT techniques. Spray and Reckase (1996) compared SB and SPRT procedures in a simulation study in terms of classification accuracy and efficiency. The results indicated that the SPRT procedure usually required fewer items than the SB procedure to achieve approximately the same level of classification accuracy. Thus, they concluded that the SPRT procedure was more efficient than the SB procedure. And this study focuses on the SPRT procedure only.

The purposes of this study are: (1) to extend the SPRT procedure to polytomous IRT models in CCT; (2) to compare polytomous items with dichotomous items using SPRT procedure for their accuracy and efficiency; (3) to study a direct approach in combining these two types of items in CCT, and (4) to study a simple method to control item exposure rate in CCT.

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 made not based on examinees' ability but based on the ratio of the likelihood of the response data conditioned at two alternative points ($\theta_0$ and $\theta_1$) around the cut-point ($\theta_c$) on the θ scale. Reckase (1983) adopted the same decision criterion to CCT even the probability model became a compound binomial model. Reckase (1983) and Spray and Reckase (1987) found this decision criterion worked well in CCT to ensure the acceptable error rate with dichotomous items.

With the increasing popularity of performance assessment in many assessment programs including certification and licensure testing, it is natural to ask if the SPRT procedure can be extended to tests that contain polytomously scored items. Few if any research has investigated how to apply polytomous items in computerized adaptive test (CAT) because of the difficulty of item scoring. Bennett, Steffen, Singley, Morley, & Jacquemin (1997) successfully adopted and scored open-ended format items in CAT that implies the feasibility of polytomous scoring in CCT in the future.

There are various polytomous IRT models have been proposed. Among them, Samejima's (1969) graded response model, Masters' (1982) partial credit model, and Muraki's (1992) generalized partial credit model (GPCM) have been widely applied to performance assessment. This study extends the SPRT procedure to the polytomous IRT models, particularly the generalized partial credit models. Under GPCM, the probability of getting a response category h 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 $\theta$ 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. As reviewed previously, this mastery decision criterion is derived originally using the binomial model and was adopted without much problem with the compound binomial model. In polytomous SPRT, the probability model is the compound multinomial model. The SPRT procedure applied to GPCM is described below.

In SPRT, 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 $\theta_j$ is an unknown parameter, and $\theta_0$ and $\theta_1$ are the lower and upper points around the cut-point ($\theta_c$). The region between $\theta_0$ and $\theta_1$ is called the indifference region.

Conditioned at these two points, we have $\pi(\theta_1)$ and $\pi(\theta_0)$, where $\pi(\theta_j)$ = Probability ($X = x \mid \theta = \theta_j$), x is the response vector. The functions, $\pi(\theta_1)$ and $\pi(\theta_0)$, are called likelihood functions of x, and a ratio of these two functions, $L(x) = \pi(\theta_1)/\pi(\theta_0)$, is called the 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 $B = \beta / (1-\alpha)$,

where $\alpha$ and $\beta$ are the error probabilities defined as follows:

Probability(choosing $H_1 \mid H_0$ is true) = $\alpha$ (false positive), and Probability(choosing $H_0 \mid H_1$ is true) = $\beta$ (false negative).

The likelihood ratio is compared to A and B to make decisions. If $L \geq A$, then $H_1$ 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.

Performance assessment with constructed response items is usually costly in testing time and scoring. It is important to inquire whether CCT with polytomous items can achieve better efficiency than CCT with dichotomous items so that much fewer polytomous items are required to achieve the same classification accuracy. Another important issue is how to combine dichotomous and polytomous items in CCT because the two types of items are often used together in real testing to enhance reliability and validity. Combining these two types of items has been studied in other testing settings (e.g., Thissen, Wainer, & Wang, 1994; Wilson & Wang, 1995; Ercikan, Burket, Julian, Link, Schwarz, & Weber, 1996; Roudabush, Candel, & Peters, 1992). Their focuses, however, were on obtaining a final score based on different types of items and they usually used some weighting scheme to derive the final score. With SPRT in CCT, the critical issue is how to combine them in achieving an optimal classification decision. A direct approach to combine these two types of items was explored in this study. That is, the likelihood ratio was computed based on the likelihood functions with the response data were directly mixed together. Simulations based on real item pools with both types of items were carried out to study the effectiveness of this direct approach.

The last issue to explore was item exposure rate control. In CCT, items are usually selected according to the information at the cutting thetas with SPRT procedure, which may induce item exposure problem: some of the items in the pool are over-used while others are under-used. In this study, a simple way to control item exposure rate was tried out. Three different item selection procedures were compared. Items were selected based on random selection, and based on a combination of item information and random selection in addition to based on “pure” item information.

## Methods

Monte Carlo simulation technique was adopted to verify the decision criterion. Simulation was also used to compare the efficiency of the polytomous items-based and dichotomous items-based SPRT procedures and the combining of the two item types. In CAT, the range of the test length is usually preset in order to cover the test content specifications on the one hand and control the item exposure rate on the other. In CCT, there are tests with different difficulty levels. In order to make the simulation more realistic, test length constraint and test difficulty were considered. Several independent variables were manipulated included:

1. Test length constraint (That is, the examinees must respond to a minimum number of items and not exceed a maximum number of items.):
   (1) minimum = 10, maximum = 50.
   (2) minimum = 1, maximum = 100.
2. Location of cutting theta (test difficulty):

(1) $\theta_c = 0.5$.

(2) $\theta_c = 1.5$.

3. Item type used:

(1) dichotomous items.

(2) polytomous items.

(3) combined items (dichotomous + polytomous).

4. Item selection based on:

(1) item information. (That is, items are selected to maximize information at the cutting theta.)

(2) information + random. (That is, a number of most informative items are selected as a group. Then, items are randomly administered to the examinee without replacement from this group. The number of items in this group depends on the test length constraint.)

(3) random. (That is, items are administered randomly without replacement.)

5. Item type ratio (dichotomous : polytomous) (for combined item pool only):

(1) 3:1. (That is, for every 4 items, 3 dichotomous items are administered first and then 1 polytomous item is administered.)

(1) 1:1. (That is, 1 dichotomous item and 1 polytomous item are administered alternatively.)

The dependent variables were: (1) classification accuracy in terms of false positive, false negative, and total error rate, and (2) number of items used to make mastery decision.

## Data

Item parameters from the 1996 NAEP Science assessment were used from the item pool. Combining three grades (4th, 8th and 12th) together, the assessment consists of 246 dichotomous items and 266 polytomous items for the study. The two types of items were calibrated on the same scale and the items for the three grades were also linked to the same scale. Item response data was generated and CCT was simulated on computer.

## Results

### Dichotomous and Polytomous Item Pools

Table 1 summarizes the results using polytomous and dichotomous item pools with different cutting thetas, length constraints, and item selection methods.

Dichotomous versus polytomous item pools. Using different cutting thetas, length constraints, and item selection methods, polytomous item pool consistently achieved better classification accuracy and uses fewer items than dichotomous item pool. For polytomous item pool, the average type I, type II, total error, and ANI rates were .016, .013, .028, and 15.127. For dichotomous item pool, the average type I, type II, total error, and ANI rates were .023, .016, .038, and 22.509.

Item selection method. Three kinds of item selection methods were applied: (a) based on item information (I), (b) based on item information and then random selection (IR), and (c) totally random selection (R).

In polytomous item pool, the average type I, type II, total error, and ANI rates were .014, .011, .025, and 9.878 based on I; were .015, .012, .026, and 13.492 based on IR; and were .018, .016, .033, and 22.012 based on R. In dichotomous item pool, the average type I, type II, total error, and ANI rates were .019, .014, .033, and 13.991 based on I; were .017, .015, .031, and 20.469 based on IR; and were .031, .019, .050, and 33.171 based on R.

These results were consistent for both dichotomous and polytomous item pool. In terms of the classification accuracy and efficiency, item selection based on I performed the best among the three methods while IR the second and R the third. The error rates (type I, type II, & total error) based on I and IR were similar either for the dichotomous or polytomous item pool. However, when item selection was based on R, both error rates and ANI were obviously higher than that of I and IR.

Test Length Constraint. Two test length constraints were applied: (a) min=10 & max=50, and (b) min=1 & max=100. In polytomous item pool, the average type I, type II, total error, and ANI rates were .017, .014, .031, and 15.459 for the constraint condition (min=10, max=50) and .014, .011, .025, and 14.795 for the constraint condition (min=1, max=100) respectively. In dichotomous item pool, the average type I, type II, total error, and ANI rates were .026, .018,

.044, and 20.457 for the constraint condition (min=10, max=50) and .019, .014, .032, and 24.560 for the constraint condition (min=1, max=100) respectively.

A clear pattern could be found for both dichotomous and polytomous item pools: the average type I, type II, total error, and ANI rates were higher for the constraint (min=10, max=50) than for the constrained (min=1, max=100). The differences of accuracy and efficiency between these two test length constrains were found more obvious for the dichotomous item pools than that of polytomous item pools. In other words, the impact of test length constraint seemed smaller to polytomous items.

Level of cutting theta. Two levels of cutting theta were applied: (a) $\theta_c$=0.5, and (b) $\theta_c$=1.5. In polytomous item pool, the average type I, type II, total error, and ANI rates were .021, .019, .04, and 17.831 for the cutting theta=.5 and were .010, .006, .016, and 12.423 for the cutting theta=1.5. In dichotomous item pool, the average type I, type II, total error, and ANI rates were .031, .025, .056, and 27.230 for $\theta_c$=.5 and .014, .007, .021, and 17.857 for $\theta_c$=1.5. Across all other criteria, as the cutting theta level increased, the type I, type II, total error, and ANI rates were found consistently decreased for both polytomous and dichotomous item pool.

## Combined Item Pool

With different item type ratios, cutting thetas, length constraints, and item selection methods, Table 2 summarizes the results of the combined item pool. This pool contained 246 dichotomous items and 266 polytomous items.

Item selection method. Similar results as dichotomous and polytomous item pools were found. Item selection according to information (I) again yielded the best classification accuracy and efficiency. Item selection according to information plus random (IR) and the random (R) were the second and the third respectively. The average type I, type II, total error, and ANI rates were .014, .011, .025, and 10.410 based on I; were .016, .014, .030, and 17.185 based on IR; and were .024, .018, .042, and 27.415 based on R.

Test length constraint. In the combined item type pool, the average type I, type II, and total error rates were slightly higher for the constraint condition (min=10, max=50) than for the constrained condition (min=1, max=100) (.020 vs. .016, .016 vs. .013, .036 vs. .029). However, the average ANI was slightly lower for the constraint condition (min=10, max=50) than for the constrained condition (min=1, max=100) (17.594 vs. 19.079). This result was consistent with those of the other two item type pools.

Level of cut-score. As the cutting theta level increased, the type I, type II, and total error rate and the ANI value were found consistently decreased. The average type I, type II, total error, and ANI rates were .025, .022, .047, and 21.972 for $\theta_c$=.5 and .011, .007, .018, and 14.701 for $\theta_c$=1.5. This result was again consistent with those of the other two item type pools.

Ratios of dichotomous to polytomous items administered. The ratios of dichotomous to polytomous items, which are 3:1 and 1:1, produced very similar results. The former ratio was just slightly lower in accuracy (type I, type II, and total error rates) and efficiency (ANI) (.019 vs. .017, vs. .015 vs. .014, .034 vs. .031, 19.344 vs. 17.329).

## Conclusion

Generally, the results of using the three item pools (polytomous, dichotomous, and combined) were consistent: item selected according to item information at the cutting theta resulted in the best classification accuracy and efficiency; test length constraint (min=1, max=100) achieved lower error and ANI rates than test length constraint (min=10, max=50); and cutting theta ($\theta_c$)=1.5 gained better results than $\theta_c$=0.5.

According to the results of this study, polytomous items work well with SPRT procedure in CCT. It was found that polytomous item pool gained more classification accuracy and utilized fewer items than dichotomous item pool. So it is concluded that polytomous items can be applied in SPRT procedure. Also, polytomous items can make more accurate mastery decision than dichotomous items with SPRT supposed other conditions being equal. However, as item consumption of polytomous item pool was not dramatically reduced in mastery decision making compared with dichotomous item pool, it cannot be concluded that polytomous item type is more efficient than dichotomous item type because the former usually takes more time to respond than the latter.

SPRT procedure was again found a good procedure for mastery decision making. With different item pools, cutting thetas, length constraints, and item selection mechanism, SPRT yielded reasonable classification accuracy and efficiency. It is confirmed once more that using item information at the cutting thetas really helps classification accuracy and efficiency in SPRT.

This study also explored a way to control item exposure rates in the content of CCT. If items are administered only according to item information, best classification accuracy and efficiency is gained. These items, however, will be over exposed. If items are totally randomly administered, the utilization rate for each item in the pool is equal yet classification accuracy and efficiency will be sacrificed. Item administration according to item information plus random (IR), in addition to test length constraint might offer a possible solution. Besides, it was found that even when item administration was in a totally random manner, the classification accuracy was still acceptable with SPRT procedure in most of the cases.

There are different ways to combine different types of items. This study explored one direct way to combine different item types by item type ratio. The results suggest that it is feasible to do so. As content balance was not taken into consideration in this study, item type ratio plus content balance could be the topic in the future research.

# 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.

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.

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.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

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.

Roudabush, G. E., Candel, G., & Peters, R. G. (1992). *Scaling constructed response items with multiple-choice items from standardized tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *No. 17.*

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.

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.

Thissen, D., Wainer, H., & Wang, X. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analyses of two tests. *Journal of Educational Measurement, 31*, 113-123.

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

Wilson, M., & Wang, W. (1995). Complex composites: Issues that arise in combining different modes of assessment. *Applied Psychological Measurement, 19*, 51-71.

Table 1. Polytomous and Dichotomous Item Pools: Error Rates & ANI

| Item Type | $\theta_c$ | Constraint | Item Selection | Type I | Type II | Total Error | ANI |
|---|---|---|---|---|---|---|---|
| polytomous | 0.5 | 10,50 | information | 0.021 | 0.017 | 0.038 | 13.315 |
| polytomous | 0.5 | 10,50 | information + random | 0.021 | 0.019 | 0.039 | 14.961 |
| polytomous | 0.5 | 10,50 | random | 0.028 | 0.028 | 0.056 | 24.017 |
| polytomous | 1.5 | 10,50 | information | 0.009 | 0.006 | 0.014 | 11.355 |
| polytomous | 1.5 | 10,50 | information + random | 0.010 | 0.007 | 0.016 | 12.010 |
| polytomous | 1.5 | 10,50 | random | 0.014 | 0.008 | 0.022 | 17.098 |
| polytomous | 0.5 | 1,100 | information | 0.018 | 0.015 | 0.033 | 9.055 |
| polytomous | 0.5 | 1,100 | information + random | 0.019 | 0.016 | 0.035 | 16.833 |
| polytomous | 0.5 | 1,100 | random | 0.020 | 0.020 | 0.040 | 28.807 |
| polytomous | 1.5 | 1,100 | information | 0.009 | 0.005 | 0.014 | 5.786 |
| polytomous | 1.5 | 1,100 | information + random | 0.009 | 0.005 | 0.014 | 10.163 |
| polytomous | 1.5 | 1,100 | random | 0.009 | 0.006 | 0.015 | 18.127 |
| dichotomous | 0.5 | 10,50 | information | 0.028 | 0.022 | 0.050 | 16.749 |
| dichotomous | 0.5 | 10,50 | information + random | 0.025 | 0.027 | 0.051 | 20.481 |
| dichotomous | 0.5 | 10,50 | random | 0.055 | 0.035 | 0.091 | 34.817 |
| dichotomous | 1.5 | 10,50 | information | 0.013 | 0.009 | 0.021 | 12.897 |
| dichotomous | 1.5 | 10,50 | information + random | 0.011 | 0.007 | 0.018 | 14.687 |
| dichotomous | 1.5 | 10,50 | random | 0.023 | 0.010 | 0.033 | 23.532 |
| dichotomous | 0.5 | 1,100 | information | 0.025 | 0.019 | 0.044 | 16.805 |
| dichotomous | 0.5 | 1,100 | information + random | 0.020 | 0.020 | 0.040 | 29.023 |
| dichotomous | 0.5 | 1,100 | random | 0.030 | 0.024 | 0.054 | 45.505 |
| dichotomous | 1.5 | 1,100 | information | 0.011 | 0.006 | 0.017 | 9.512 |
| dichotomous | 1.5 | 1,100 | information + random | 0.010 | 0.006 | 0.016 | 17.686 |
| dichotomous | 1.5 | 1,100 | random | 0.015 | 0.008 | 0.023 | 28.830 |

Note: $\theta_c$ is the cutting theta. Constraint is the test length constraint. Item selection is the way to select items to administer. Type I is false positive error. Type II is false negative error. Total error is Type I plus Type II. ANI is the average number of items used.

Table 2. Combined Item Pools: Error Rates & ANI

| **Item Type** | **Ratio** | $\theta_c$ | **Constraint** | **Item Selection** | **Type I** | **Type II** | **Total Error** | **ANI** |
|---|---|---|---|---|---|---|---|---|
| combined | 3:1 | 0.5 | 10,50 | information | 0.021 | 0.018 | 0.039 | 13.911 |
| combined | 3:1 | 0.5 | 10,50 | information + random | 0.025 | 0.023 | 0.048 | 18.795 |
| combined | 3:1 | 0.5 | 10,50 | random | 0.041 | 0.034 | 0.075 | 31.220 |
| combined | 3:1 | 1.5 | 10,50 | information | 0.009 | 0.006 | 0.015 | 11.643 |
| combined | 3:1 | 1.5 | 10,50 | information + random | 0.010 | 0.007 | 0.017 | 13.478 |
| combined | 3:1 | 1.5 | 10,50 | random | 0.022 | 0.011 | 0.033 | 21.857 |
| combined | 3:1 | 0.5 | 1,100 | information | 0.020 | 0.015 | 0.035 | 10.898 |
| combined | 3:1 | 0.5 | 1,100 | information + random | 0.021 | 0.021 | 0.042 | 25.591 |
| combined | 3:1 | 0.5 | 1,100 | random | 0.028 | 0.025 | 0.053 | 39.568 |
| combined | 3:1 | 1.5 | 1,100 | information | 0.009 | 0.005 | 0.014 | 6.991 |
| combined | 3:1 | 1.5 | 1,100 | information + random | 0.008 | 0.007 | 0.015 | 14.824 |
| combined | 3:1 | 1.5 | 1,100 | random | 0.012 | 0.007 | 0.019 | 23.356 |
| combined | 1:1 | 0.5 | 10,50 | information | 0.019 | 0.020 | 0.039 | 13.175 |
| combined | 1:1 | 0.5 | 10,50 | information + random | 0.024 | 0.020 | 0.044 | 16.569 |
| combined | 1:1 | 0.5 | 10,50 | random | 0.033 | 0.027 | 0.060 | 27.681 |
| combined | 1:1 | 1.5 | 10,50 | information | 0.009 | 0.006 | 0.015 | 11.338 |
| combined | 1:1 | 1.5 | 10,50 | information + random | 0.011 | 0.007 | 0.018 | 12.954 |
| combined | 1:1 | 1.5 | 10,50 | random | 0.017 | 0.009 | 0.026 | 18.507 |
| combined | 1:1 | 0.5 | 1,100 | information | 0.019 | 0.015 | 0.034 | 9.024 |
| combined | 1:1 | 0.5 | 1,100 | information + random | 0.022 | 0.020 | 0.042 | 22.197 |
| combined | 1:1 | 0.5 | 1,100 | random | 0.025 | 0.023 | 0.048 | 35.037 |
| combined | 1:1 | 1.5 | 1,100 | information | 0.009 | 0.005 | 0.014 | 6.303 |
| combined | 1:1 | 1.5 | 1,100 | information + random | 0.009 | 0.006 | 0.015 | 13.072 |
| combined | 1:1 | 1.5 | 1,100 | random | 0.010 | 0.007 | 0.017 | 22.091 |

Note: Ratio is the proportion of dichotomous items to polytomous items administration.