Running Head: RESPONSE TIME THRESHOLD

Setting The Response Time Threshold Parameter to Differentiate Solution Behavior From Rapid-Guessing Behavior

Xiaojing J. Kong, Dennison S. Bhola, & Steven L. Wise

James Madison University

Paper presented at annual meeting of the National Council on Measurement in Education, Montreal, April, 2005.

Setting the Response Time Threshold Parameter to Differentiate Solution Behavior from Rapid-Guessing Behavior

In this study we compared four methods for setting a response time threshold that differentiates rapid-guessing behavior from solution behavior when examinees are obliged to complete a low-stakes test. The four methods examined were: (1) a fixed threshold for all test items; (2) thresholds based on item surface features such as the amount of reading required; (3) thresholds based on visually inspecting response time distributions; (4) thresholds statistically generated based on a two-state mixture model (Schnipke & Scrams, 1997, 2002). To compare the sets of threshold parameters, we used the method designed by Wise and Kong (2005) to assess the reliability and validity of response time effort scores, which were generated on the basis of the specified threshold values. Our results showed only minor differences among the four threshold identification methods. Recommendations were given regarding the uses of the methods.

The use of response time to detect examinees' test-taking behavior has long been attractive to researchers. Schnipke and Scrams (e.g., Schnipke & Scrams, 1997, 2002) studied response times in the context of speeded high-stakes tests. In their studies, examinees were observed to engage in either *solution behavior* or *rapid-guessing behavior* on each item. Examinees exhibit solution behavior when they actively seek to correctly answer test items. As time is running out, however, examinees exhibit rapid-guessing behavior by responding so quickly that they could not have had enough time to fully consider the item. These two types of test-taking behaviors have also been observed under unspeeded, low-stakes testing conditions (Wise & Kong, 2005). Unlike Schnipke and Scrams, however, Wise and Kong believed that rapid-guessing behavior during a low-stakes test indicated a lack of examinee motivation and low effort rather than the hurrying-to-finish strategy employed by examinees in a high-stakes context especially when there is no penalty for guessing. That is, in a low-stakes testing environment, examinees who are willing to give good effort to a test item will be very likely to exhibit solution behavior, while those who are not motivated will be likely to put forth little effort and exhibit rapid-guessing behavior.

In Wise and Kong's (2005) study, item response time proved useful in detecting the effort given by examinees in low-stakes computer-based tests (CBTs), and rapid guessing was found throughout a test session, not just toward the end. Conceptually, a test session consists of a series of examinee-item encounters, in which either solution or rapid-guessing behavior is identified based upon the time the examinee spends on the item. Thus, for an item i, there is a threshold, T_i, that represents the response time boundary between rapid-guessing behavior and solution behavior. Given an examinee j's response time, RT_{ij}, to item i, a dichotomous index of item solution behavior, SB_{ij}, is computed as

$$SB_{ij} = \begin{cases} 1 & \text{if } RT_{ij} \geq T_i, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

The index *SB* is crucial to several applications. Wise and Kong (2005) used SB as a basic component of a new measure of examinee test-taking effort, which they termed response time effort (RTE). The index of overall response time effort for examinee j to the test is given by

$$RTE_j = \frac{\sum SB_{ij}}{k}, \quad (2)$$

where k = the number of items in the test.

An RTE score represents the proportion of test items for which a specific examinee exhibited solution behavior. RTE scores range from zero to one, with the larger values indicating stronger examinee effort during the test. Wise and Kong (2005) investigated the psychometric characteristics of RTE scores, and found adequate evidence for score reliability and validity that supported their claim for using RTE scores as a measure of test-taking effort.

The index *SB* has also been employed in Wise (2004)'s investigations of the differential effort received by individual test items. He found that the strongest predictors of the effort received by items were item length and item position (Wise, 2004). It was also found that psychometric properties of the test were positively affected by treating responses arising from rapid guesses as missing.

Another important application of *SB* can be seen in Wise and DeMars (in press)'s research on an improved IRT model. This model was termed an effort-moderated IRT model. According to their research findings, the effort-moderated model performed better than the standard 3PL model with respect to model fit, the accuracy of item parameter and test information estimates, as well as the validity of proficiency estimates (Wise & DeMars, in press).

It is evident that the SB index is of great significance to various applications. The initial and crucial task in obtaining an accurate SB value is setting the time threshold. Conceptually, the threshold for a given item, T_i, as shown in Equation (1), defines the range of response times that would be too short for an examinee to have a reasonable chance to read the item and identify the correct answer. In practice, we want the threshold to differentiate solution behavior from rapid-guessing behavior as accurately as possible. Schnipke and Scrams (1997) used a two-state mixture model approach to set the thresholds, which were specified as the points where the distributions of rapid-guessing and solution behavior crossed for each item. Wise and Kong (2005) assigned different thresholds for different items according to the amount of reading/scanning the item required. A third method for identifying threshold is empirical. Wise and DeMars (in press) primarily used this method in their study of effort-moderated IRT model. More specifically, the shape of response time distributions were found to be similar – the distributions are unimodal and positively skewed with a small frequency spike occurring during an initial short time period. However, the "width" of the spike varied substantially across items, extending from only a few seconds up to over 10 seconds. By visually

inspecting the response time distribution, Wise and DeMars defined the threshold for a given item as the end point of the short time spike (Wise & DeMars, in press). Figure 1 shows the response time frequency distribution and the identified threshold position for one achievement test item. In addition to these methods, a relatively simple way to define a threshold is to set a fixed threshold for an entire test. In so doing, we assume that a common threshold, instead of varying thresholds, can perform well enough for all the items in the test.

The purpose of this study was to compare the effectiveness of four different methods by which thresholds could be specified. The four methods examined were: (1) a fixed threshold for all items; (2) thresholds for each item based on two surface features of the items: item length and whether a particular ancillary reading is provided for the first time; (3) thresholds for each item established by visually inspecting each item's response time distribution; and (4) thresholds for each item generated statistically using a two-state mixture model (Schnipke & Scrams, 1997, 2002).

Method

Participants

Participants in this study were 524 undergraduate students (55.2% females and 44.8% males) at a mid-sized, southeastern state University who were required to participate in a University-wide Assessment Day during the spring of 2004. Students with 45-70 credit hours were mandated to participate in assessment. The purpose of testing was to determine how well the University was doing in educating students. Since there were no personal consequences associated with students' performance, this can be considered a low-stakes testing environment.

To assess each participant's level of academic ability, Scholastic Assessment Test (SAT) scores were obtained from the student records database. One or more SAT scores were missing for 36 examinees; these examinees were consequently deleted from the analyses, resulting in a final sample size of 488.

Measures

Achievement Test. The achievement test used in this study was a 60-item computerized Information Literacy Test (ILT). This locally developed multiple-choice test was designed to match the Association of College and Research Libraries information literacy competency standards. It assesses student skills in finding, critically evaluating, and effectively using information. On many of the ILT items, examinees were provided ancillary online materials, such as tables, figures, or websites for them to read/study in order to answer the question posed in the item. The number of response options per item ranged from two to five. ILT scores represent the percentage of items the examinee answered correctly. In the current study, the reliability of the ILT scores was acceptably high, with coefficient alpha equal to .87.

Student Opinion Survey (SOS). The SOS (Sundre, 1999; Wolf & Smith, 1995) is a 10-item motivation scale that yields a total score and two 5-item subscale scores (Reported Effort, Perceived Importance of the Test). The Reported Effort subscale of the SOS was used to measure examinee self-reported effort on the ILT. Each SOS item uses a five-point response scale ranging from strongly disagree to strongly agree. Research has demonstrated adequate reliability for each subscale score and the total score as well as substantial favorable validity evidence (Sundre, 1999; Sundre & Kitsantas, 2004; Sundre & Moore, 2002). In this study, a computer-based version of SOS was administered.

Procedure

Participants were randomly assigned to testing rooms based on the last two-digits of their student identification number. Before the tests started, trained university proctors explained to the participants the value of the assessment data to the university. Participants were not given feedback on their test performance, either during or after the test.

Testing was conducted in several university computer laboratories. The examinees were administered the ILT first, and then the SOS. Although a time limit of 90 minutes was imposed during administration of the ILT, 100% of the examinees finished within an hour. Thus, the ILT could be considered an unspeeded test. The SOS took approximately two additional minutes to complete.

Threshold Types

Method 1. A Fixed Threshold (3SEC). Given the content area measured by the ILT and the amount of reading/scanning that the items required, a three-second threshold was used for all items. Practically, the approach of a fixed threshold for all items was the easiest one to implement.

Method 2. Item Feature Rule-Based Thresholds (READ). Visual inspection of the items and their characteristics suggested that item response times could be strongly related to item features. Therefore, one threshold was specified for each item based on two surface features of items: item length (quantified as number of characters) and whether a particular ancillary reading is provided for the first time. More specifically, if an item was shorter than 200 characters, a 3-second threshold was used. If an item was longer than 1000 characters, or if the item provided some particular ancillary reading for the first time, a 10-second threshold was used. For the remaining items, a 5-second threshold was used.

Method 3. Empirical Identification of Thresholds (INSPECT). By examining the response time distributions of all items, we found that the distribution shapes were similar among items. Consistent with previous research findings, the item response time distributions seemed unimodal and positively skewed, with an additional small frequency spike for very short response times. This shape of response time distribution with a frequency spike occurring long before the median response time, was observed regardless

of the magnitude of the threshold. For each item, four evaluators, including two faculty members and two graduate students in the Assessment and Measurement doctoral program, were asked to identify the threshold response time based upon their visual inspection of the data and expert judgment. A mean value was used as the final threshold value for each item based upon this method.

Method 4. Mixture-Model-Based Thresholds (MIXTURE). Each test item was reviewed, and a hypothesis was made that the item should have a bimodal (two component) response time distribution. The hypothesis was consistent with previous research by Schnipke & Scrams (1997, 2002). Procedurally, a normal kernel density plot of the response times was produced, and starting values for estimation of the normal mixture model were given based upon visual inspection of the plots. For each item, response time distribution was fit with kernel smoothing (Wand & Jones, 1995; van Zandt, 2000) and finite mixture models (McLachlan & Peel, 2000). Parameters of the best-fitting model obtained were used to generate the thresholds.

Data Analysis

To compare the sets of threshold parameters, we used the method designed by Wise and Kong (2005) to evaluate the reliability and validity of RTE scores. More specifically, (1) the reliability RTE scores was examined. Cronbach's coefficient alpha was used, with an internal consistency value of .80 judged to be the minimal criterion for adequate reliability; (2) evidence regarding the convergent validity of RTE scores was evaluated using the correlation of RTE scores with self-reports of effort obtained from the SOS-Effort subscale; (3) evidence regarding the discriminant validity of RTE scores was evaluated using the correlation of RTE scores with Scholastic Achievement Test (SAT) scores; (4) item response accuracy rates under solution behavior and rapid-guessing behavior were compared with the accuracy rate based on chance; (5) motivation filtering effects were compared among the sets of the threshold parameters. In motivation filtering, the data from examinees exhibiting low test-taking effort on an achievement test are removed, or filtered, from the dataset. Previous research has shown that when a motivation filtering technique was used, test performance improved; test score reliability remained relatively constant; and the correlation between test performance and an external variable showed a substantial increase (Sundre & Wise, 2003; Wise & DeMars, in press; Wise & Kong, in press; Wise, V.L., Bhola &Wise, S. L., 2005).

Results and Discussion

The primary purpose of threshold identification is to set the response time boundary between rapid-guessing behavior and solution behavior. As described previously, four sets of thresholds resulting from different identification methods were evaluated in the current study. Correspondingly, four sets of RTE scores were computed, using Equation (2). Figure 2 shows the positively skewed distribution of RTE scores using the READ threshold. RTE distributions were found to be almost identical for all the threshold types.

Descriptive statistics for RTE scores were shown in Table 1 to provide a basis of comparison among the four threshold types. As we can see, across the threshold types, the mean RTE score was consistently in the mid-.90s. This value is high and indicates that, regardless of the threshold setting method used, the data indicates that, on average, examinees engaged in solution behavior on over 90% of the items on the ILT. The standard deviation of the RTE scores was also similar in magnitude (.16-.18) and consistent across methods, and indicated that the spread of RTE scores about the mean was about the same regardless of threshold setting method used. In addition, our results also showed that the internal consistency reliability of RTE scores was very high and identical across all four threshold identification methods.

Table 2 presents validity evidence for the RTE scores resulting from each threshold type. Correlations between RTE scores and performance on the ILT test performance were large (consistently above .7 for all of the methods), and statistically significant across methods, with a slightly lower correlation coefficient when thresholds were set using the 3SEC method.

RTE scores were also correlated with self-reported effort scores. Since both of these are measures of students' test-taking effort, it was expected that the correlation between RTE scores and self-reported effort scores would be high. The observed correlations between RTE scores and self-reported effort scores were of moderate magnitude (around .40) and statistically significant with thresholds set using all four methods. Again, a slightly lower correlation coefficient was observed when thresholds were set using the 3SEC method. The same tendency can also be observed in the correlations of RTE scores with total time spent on the test.

The third criterion for evaluating threshold types pertains to discriminant validity evidence. As displayed in Table 2, correlations between RTE and SAT scores were close to zero, indicating that RTE scores were not related to examinees' academic ability. These findings were consistent across threshold types.

The fourth criterion compared item response accuracy rates under solution behavior and rapid-guessing behavior. The underlying assumption is that solution behavior yields item scores whose accuracy exceeds chance while the accuracy of rapid-guessing responses is near chance level. Table 3 indicates that the accuracy rates of rapid-guessing responses were far below those of solution responses, and did not significantly exceed chance level. There were no substantial differences among the results generated based upon the different threshold types. However, INSPECT and MIXTURE yielded rapid-guessing accuracy that most resembled chance.

Finally, we examined the motivation filtering effects using the four sets of RTE scores (see Table 4). It was observed that when motivation filtering was done using RTE scores, (a) test performance improved, (b) standard error of measurement remained relatively constant, and (c) the correlation between test performance and SAT verbal scores increased. All of the threshold types performed similarly, with the 3SEC method faring slightly less than the other methods.

Conclusions and Recommendations

The four threshold identification methods investigated in this study gave very comparable results in terms of reliability and validity evidence. This suggests that response time effort is not highly sensitive to which threshold identification method is used.

In practice, the information needed to implement the four methods is different. MIXTURE is a purely statistical method. It requires substantial response time data and expertise in computation. When sufficient resources are available, it is a viable option that does not depend much on human subjectivity. However, generating and interpreting threshold can be difficult due to procedural and computational complexity. The INSPECT method also requires response time data. But visual inspection of response time distributions is a relatively easy process and does not require computation expertise. To implement the READ identification method, we can simply use the item surface features such as the amount of reading demanded by the items. A distinguishing feature of this method is that it could be used before response time data is collected. We recommend this method when response time data is not available. This method still allows the identification of varying thresholds for individual items. The method of using a fixed threshold for all items is the simplest. It is easy to program and can be conveniently used without data. However, the upfront work needed for determining the single threshold requires close examination of the content and surface features of all items and choose a single threshold that would be suitable for use across the entire test. It is recommended that this method be used when one does not have prior response time data for the items, and when all of the items are similar in terms of content, format and cognitive complexity.

In conclusion, all the four threshold identification methods perform similarly. The choice of the most appropriate method is largely dependent on the resources available, such as stored response time data. It is comforting that the three methods allowing varying thresholds yielded nearly identical threshold values. Such a convergence of results across very different methods encourages confidence in the threshold identification process.

References

Bergstrom, B. A., Gershon, R. C., & Lunz, M.E. (1994, April). *Computer adaptive testing: Exploring examinee response time using hierarchical linear modeling.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Bhola, D. S. (1994). *An investigation to determine whether an algorithm based on response latencies and number of words can be used in a prescribed manner to reduce measurement error*. Unpublished doctoral dissertation, University of Nebraska-Lincoln.

Bhola, D. S., Plake, B.S., & Roos, L. L. (1993, October). *Setting an optimum time limit for a computer-administered test.* Paper presented at the annual meeting of the Midwestern Education Research Association, Chicago, IL.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B, 39*, 1-38.

Gershon, R. C., Bergstrom, B. A., & Lunz, M. (1993, April). *Computer adaptive testing: Exploring examinee response time.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Halkitis, P. N., & Jones, J. P. (1996). *Estimating testing time: The effects of item characteristics on response latency.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Kingsbury, G.G., Zara, A. R., & Houser, R. L. (1993, April). *Procedures for using response latencies to identify unusual test performance in computerized adaptive tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Kingsbury, G.G., Zara, A. R., & Houser, R. L. (1994, April). *Modeling item response latencies in computerized adaptive tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

McLachlan, G. & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Reese, C. M. (1993, April). *Establishing time limits for the GRE computer adaptive tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Schnipke, D. L. (1995, April). *Assessing speededness in computer-based tests using item response times.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Schnipke, D. L. (1996, April). *How contaminated by guessing are item-parameter estimates and what can be done about it?* Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY. (ERIC Document Reproduction Service No. ED400276)

Schnipke, D. L. (1999). *The influence of speededness on item-parameter estimation* (Computerized Testing Report No. 96-07). Princeton, NJ: Law School Admission Council. (ERIC Document Reproduction Service No. ED467809)

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement, 34*, 213-232.

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. In Mills, C. N., Potenza, M.T., Fremer, J. J., & Ward, W. C. (Eds.). *Computer-based testing: Building the foundation for future assessments.* Mahwah, NJ: Lawrence Erlbaum Associates.

Sundre, D. L. (1999, April). *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service No. ED432588)

Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14* (1), 8-9.

Sundre, D. L., & Wise, S. L. (2003, April). *'Motivation filtering': An exploration of the impact of low examinee motivation on the psychometric quality of tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Thissen, D. (1983). Timed testing: An approach using item response testing. In D.J. Weiss (Ed.), *New horizons in testing: Latent trait theory and computerized adaptive testing* (pp. 179-203). New York: Academic Press.

van der Linden, W. J., Scrams, D. J., & Schnipke, D. L. (1998). *Using response-time constraints in item selection to control for differential speededness in computerized apaptive testing* (Report No. 98-06). Enschede, Netherlands: Twente University.

van Zandt, T. (2000). How to fit a response time distribution. *Psychometric Bulletin & Review, 7*, 434-465.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman and Hall.

Wise, S. L. (2004). *An investigation of the differential effort received by items on a low-stakes, computer-based test.* Manuscript submitted for publication.

Wise, S. L., & DeMars, C. E. (2005). Examinee motivation in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*, 1-17.

Wise, S. L., & DeMars, C. E. (in press). An application of item response time: the effort-moderated IRT model. *Journal of Educational Measurement.*

Wise, S. L., & Kong, X. (in press). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education.*

Wise, V. L., Bhola, D. S., & Wise, S. L. (2005). *The generalizability of motivation filtering in improving test score validity.* Manuscript submitted for publication.

Acknowledgement

The authors wish to thank Dr. Patrick Meyer for his kind and valuable help with the data analyses.

TABLE 1

Descriptive Statistics for Response Time Effort (RTE) Based on Threshold Types

Threshold Type	Mean	Standard Deviation	Median	Coefficient Alpha
3SEC	.95	.16	1.00	.99
READ	.93	.18	1.00	.99
INSPECT	.94	.18	1.00	.99
MIXTURE	.94	.18	1.00	.99

TABLE 2

Correlations of Response Time Effort with Other Variables for Each Threshold Type

	Threshold Type			
Measure	3SEC	READ	INSPECT	MIXTURE
1. Test Performance	.741*	.775*	.772*	.774*
2. Self-Reported Effort	.381*	.411*	.402*	.402*
3. Total Test Time	.615*	.662*	.648*	.648*
4. SAT-Verbal	.073	.076	.081	.078
5. SAT-Quantitative	-.042	-.055	-.048	-.049

Note. $N = 488$

*$p < .01$

TABLE 3

Item Response Accuracy Relative to Random Responding for Rapid-Guessing Responses and Solution Responses on 60 ILT items

	Solution Responses					Rapid-Guessing Responses				
Threshold Type	M	SD	*t*	*p value*	*Effect Size*	M	SD	*t*	*p value*	*Effect Size*
3SEC	71.74	22.72	15.93	<.001	2.05	27.48	17.72	1.08	.29	.14
READ	72.18	22.75	16.06	<.001	2.07	29.24	17.97	1.82	.07	.23
INSPECT	72.21	22.74	16.07	<.001	2.08	25.84	14.90	.43	.67	.06
MIXTURE	72.26	22.74	16.09	<.001	2.08	26.11	14.63	.58	.57	.07

Note. Scores are expressed as percentages. Test Performance scores had an expected accuracy value under random responding of 25.02.

The effect size index was given by $(M_{obtained} - M_{chance})/SD_{obtained}$.

TABLE 4

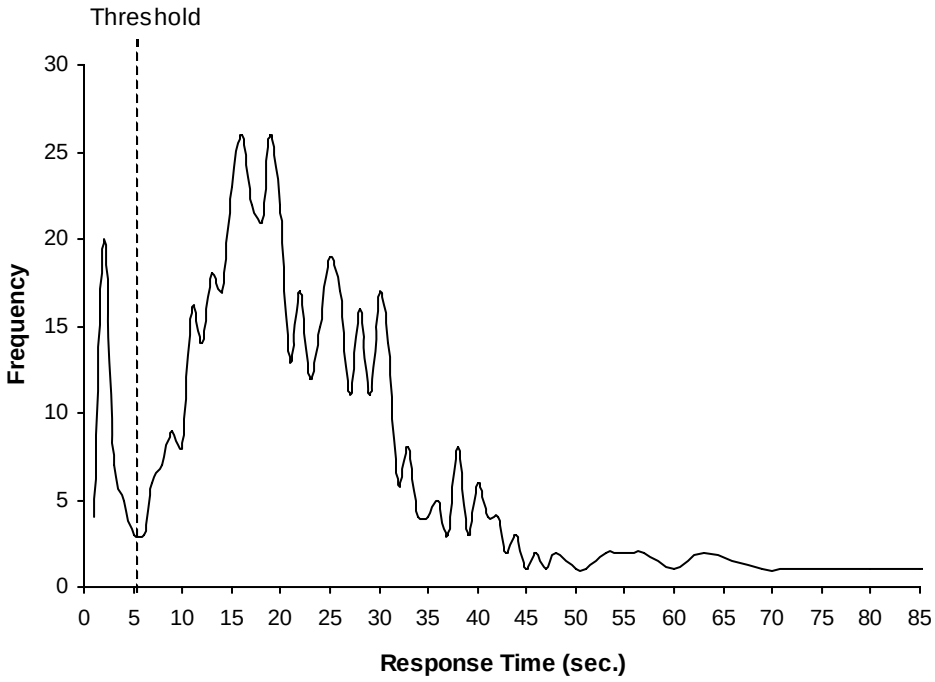
The Effects of Motivation Filtering for the ILT Data Using Response Time Effort as Filters

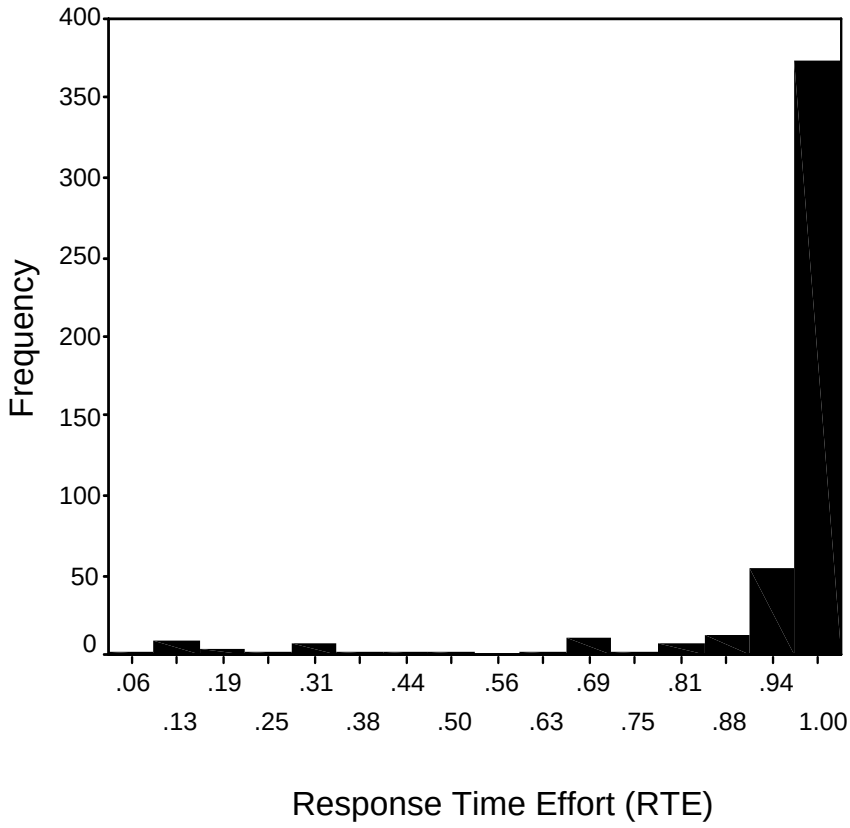
Data Analyzed	n	M	SD	SEM	Corr. Between ILT and SAT-V	Mean SAT-V Score
All Examinees	488	41.68	8.37	3.02	.36	578.69
RTE Using 3SEC Threshold						
Those with RTE ≥ .60	464	42.93	6.45	2.96	.44	579.70
Those with RTE ≥ .70	462	43.01	6.33	2.97	.45	579.65
Those with RTE ≥ .80	455	43.27	6.01	2.94	.46	580.09
Those with RTE ≥ .90	442	43.59	5.66	2.89	.45	581.36
RTE Using READ Threshold						
Those with RTE ≥ .60	461	43.06	6.26	2.94	.44	580.04
Those with RTE ≥ .70	451	43.43	5.78	2.89	.44	580.95
Those with RTE ≥ .80	445	43.60	5.60	2.91	.46	580.85
Those with RTE ≥ .90	430	43.88	5.39	2.90	.48	581.14
RTE Using INSPECT Threshold						
Those with RTE ≥ .60	461	43.06	6.26	2.94	.44	580.04
Those with RTE ≥ .70	454	43.30	5.97	2.92	.45	580.53
Those with RTE ≥ .80	447	43.54	5.66	2.89	.46	580.67
Those with RTE ≥ .90	431	43.85	5.41	2.91	.47	581.16
RTE Using MIXTURE Threshold						
Those with RTE ≥ .60	461	43.06	6.26	2.94	.44	580.04
Those with RTE ≥ .70	453	43.36	5.84	2.92	.44	580.99
Those with RTE ≥ .80	445	43.57	5.65	2.88	.46	580.54
Those with RTE ≥ .90	431	43.86	5.40	2.91	.48	581.00

Figure Captions

Figure 1. Distribution of examinee response times for an ILT item.

Figure 2. Histogram of Response Time Effort (RTE) scores based on the READ threshold.


Threshold
30
25
20
15
10
5
0
Frequency
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
Response Time (sec.)


400
350
300
250
200
150
100
50
0
Frequency
.06
.13
.19
.25
.31
.38
.44
.50
.56
.63
.69
.75
.81
.88
.94
1.00
Response Time Effort (RTE)