DOCUMENT RESUME

ED 400 287 TM 025 556

AUTHOR Bergstrom, Betty; And Others
TITLE Computerized Adaptive Testing Exploring Examinee Response Time Using Hierarchical Linear Modeling.
PUB DATE Apr 94
NOTE 26p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adaptive Testing; Adults; Certification; *Computer Assisted Testing; Difficulty Level; *Reaction Time; Responses; Student Characteristics; Test Anxiety; Test Construction; *Test Items; Test Length; *Timed Tests
IDENTIFIERS *Hierarchical Linear Modeling

ABSTRACT

Examinee response times from a computerized adaptive test taken by 204 examinees taking a certification examination were analyzed using a hierarchical linear model. Two equations were posed: a within-person model and a between-person model. Variance within persons was eight times greater than variance between persons. Several variables significantly predicted within-person variance. Response time increased with increasing items, test length, and increasing relative item difficulty. Item sequence was negatively related to response time, and some content areas required more time than others. Examinees spent more time on items they got wrong than on items they got right, and they took longer to respond when the correct answer was A, B, or C than when the correct answer was D. Only one variable, test anxiety, significantly predicted variance between examinees. Examinee age, sex, first language, and ethnicity did not predict between-person variance, and low-ability examinees did not take longer to respond to items than high-ability examinees. Understanding how item characteristics impact on response time may allow test developers to allot total test time based on the response time history of the individual test items. This study also suggests that examinee characteristics are generally not related to response time, but that more controllable factors such as item length, position of the keyed correct answer, and use of figures do contribute to response items. An appendix contains the anxiety survey. (Contains 1 figure, 5 tables, and 12 references.) (Author/SLD)

Computerized Adaptive Testing Exploring Examinee Response Time Using Hierarchical Linear Modeling

Betty Bergstrom
Richard Gershon
Computer Adaptive Technologies, Inc

Mary E. Lunz
American Society of Clinical Pathologists

Paper presented at the annual meeting
National Council on Measurement in Education
New Orleans, Louisiana

April, 1994

Abstract

Examinee response times from a computerized adaptive test were analyzed using a hierarchical linear model. Two equations were posed: a within-person model and a between-person model. Variance within persons was eight times greater than variance between persons. Several variables significantly predicted within-person variance. Response time increased with increasing item text length and increasing relative item difficulty. Item sequence was negatively related to response time and some content areas required more time than others. Examinees spent more time on items they got wrong than on items they got right, and they took longer to respond when the correct answer was A, B, or C than when the correct answer was D. Only one variable, test anxiety, significantly predicted variance between examinees. Examinee age, sex, first language and ethnicity did not predict between-person variance and low able examinees did not take longer to respond to items than high able examinees. Understanding how item characteristics impact on response time may allow test developers to allot total test time based upon the response time history of the individual test items. This study also suggests that examinee characteristics are generally not related to response time, but that more controllable factors such as item length, position of the keyed correct answer, and use of figures do contribute to response time.

Computerized Adaptive Testing
Exploring Examinee Response Time
Using Hierarchical Linear Modeling

Computerization allows previously unobtainable data, such as item response time, to be collected and used to improve tests. Researchers can now analyze the amount of time that an individual takes when answering a test item. Understanding the various factors which impact on item response time should be useful to test developers to predict the amount of time required for total test administration. Response time may also be used in the future to help identify unusual or cheating behaviors (Kingsbury, Zara and Houser, 1993). For example, an individual who responds to items more quickly than within the normal response range could be flagged for further investigation.

The arrival of computerized adaptive testing further complicates the importance of overall test time and individual item time. The archetype of computer adaptive testing would have the test taker answer only as many items as needed to determine the individual's ability relative to a pre-determined pass/fail point or within a specified level of precision. This strategy would typically not seek to place any real limits on the amount of time that an individual could spend on any single item, and total test time would vary between individuals based upon both the number of items required to determine ability and the examinee's tendency to take more or less time in answering items. However, in real testing situations, maximum time limits are usually set due to cost or other administrative issues.

Rafaeli and Tractinsky (1991) found a strong negative correlation between response time and accuracy for general knowledge tests, but not for mathematical reasoning tests administered by computer. They suggested that in adaptive survey techniques, time could be allocated to each item based upon the difficulty of the item and test taker ability. The resulting test could be significantly shorter, although corresponding examinee satisfaction ratings were likely to decline. Rafaeli and Tractinsky also reported that examinees took longer when time limitations were based on the total test than when time limits were placed on each item. Examinees, however, reported a preference for a total test time limitation.

As is the case with all new technologies, there is a general sense of fear regarding the differential impact of using a computer to administer tests versus traditional paper and pencil methods (Fair Test, 1992). Several studies have addressed the issue of overall test-taking time on test scores, and additionally whether differences in time interact with demographic variables such as race and sex. Generally, these studies found that increased overall testing time improves scores, but they found no significant interactions with race or sex (Wild, Durso, and Rubin, 1982; Evans and Reilly, 1976).

Previous regression studies of item response time on computerized adaptive tests have left much of the variation in time unexplained (Gershon, Bergstrom and Lunz, 1993; Kingsbury et al., 1993). Kingsbury, Zara and Houser (1993) reported that none of the variables they investigated accounted for more than 8% of the variance in item response times.

In our initial study (Gershon et al., 1993), we found several variables to be significant predictors of examinee response time per item. Response time increased with increasing item text length and increasing item difficulty. Examinees took longer to respond to items at the beginning of the test than at the end of the test. Response time also varied by content category, whether or not the item contained an illustration, distractor position of the correct response, and whether or not the examinee got the item correct. Item level variables accounted for 19% of the variance in response time. Examinee variables (test anxiety, gender, ethnic background, age and language) accounted for an additional 2% of the variance in response time.

This previous research treated individual examinee/item response times as independent pieces of data when, in fact, item observations are nested within persons. We believed that at least some of the unexplained variance in response time can be attributed to individual person differences.

In this study, we reanalyze our original data using a hierarchical linear model (Bryk, Raudenbush, Seltzer and Congdon, 1989) which allows us to separate individual variation in response time from error variation, giving better estimates of individual variation and permitting exploration of the effects of item level characteristics and person demographics on response time. We hypothesize that:

1) Item characteristics such as text length, sequence of the item on a test, presence of a figure, and location of the correct answer (a, b, c, d) will significantly predict variance in response time.

2) Response time variance will be greater across candidates than within candidates.

3) Response time will vary across candidates based on demographic characteristics (age, gender, test anxiety, ethnic background, etc.).

Data and Instruments

Data were collected from a certification examination administered in 1991 using a computerized adaptive algorithm. Examinees had the option of taking a computerized adaptive test or the traditional paper and pencil test. Two hundred four examinees chose to take the computerized test.

The computerized test was administered with the CAT ADMINISTRATOR software (Gershon, 1990) using the PROX method of estimation (Wright & Stone, 1979) for the item selection algorithm. A pre-calibrated item bank consisting of 696 items was prepared for the examination. Each test item fit on one computer screen, so examinees did not have to scroll item text. Content was distributed across six content areas. A content balancing mechanism, to insure that item distribution matched the test specifications of the traditional written test, was included in the computerized adaptive algorithm.

Each test began with an item randomly chosen from items within .10 logits in difficulty of the pass/fail point. The following 9 items were constrained to within .10 logits of the previously administered item difficulty. This procedure effectively constrained the difficulty of the first 10 items to within 1 logit of the pass/fail point. Thereafter, items were targeted to a 60% probability of correct response. Items were chosen at random from unused items within .10 logits of the targeted item difficulty within the specified content area.

Testing stopped when the estimated examinee ability was 1.65 times the standard error of measurement above or below the pass point. Minimum test length was 50 items and maximum test length was 100 items. Examinees were allotted 2 hours to complete the test.

Some items contained figures or color plates. These graphics were contained in a separate illustration booklet. When this type of item appeared on the screen, examinees were instructed to refer to a specific illustration in the booklet.

Upon completion of the adaptive test, examinees were allowed to review and change answers. This paper however, deals only with response time during the initial test administration. Analysis of review times is not included. (See Lunz, Bergstrom and Wright, 1992 for further information on the effect of allowing review).

Method

Response time was recorded from the time the item appeared on the screen until the examinee pressed enter and moved to the next item. Figure 1 shows the distributions of response time and the log of response time. The mean time per item was 63 seconds with a standard deviation of 47 seconds. For this analysis, the log of response time better approximates a normally distributed random variable and therefore was used as the dependant variable.

INSERT FIGURE 1 HERE

Data were analyzed with a hierarchical linear model (Bryk et al., 1989). Computations were performed with a software program for fitting hierarchical linear models (Thum, 1994). With the hierarchical linear model we pose two equations, a within-person and a between-person model. The within-person model specifies the relationships between t_{ij}, the log of the observed response time on item *i* for examinee *j* and various independent variables X_{ijp}. Thus the within-person model can be written as:

$$t_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \beta_{j3}X_{ij3} + \beta_{j4}X_{ij4} + \beta_{j5}X_{ij5} + \beta_{j6}X_{ij6} + \beta_{j7}X_{ij7} + \beta_{j8}X_{ij8} + \beta_{j9}X_{ij9} + \beta_{j10}X_{ij10} + \beta_{j11}X_{ij11} + \beta_{j12}X_{ij12} + \beta_{j13}X_{ij13} + R_{ij} \qquad (1)$$

where

t_{ij} is the log of the response time on item i for examinee j;

X_{ij1} is the relative difficulty of item i for examinee j[1];

X_{ij2} is the correctness of response of item i for examinee j (0 = wrong, 1 = right);

X_{ij3} is the position of administration (sequence) of item i for examinee j;

X_{ij4} is the item length;[2]

X_{ij5} is the content category of item i for examinee j (0= Content 6, 1= content 1);

X_{ij6} is the content category of item i for examinee j (0= Content 6, 1= content 2);

X_{ij7} is the content category of item i for examinee j (0= Content 6, 1= content 3);

X_{ij8} is the content category of item i for examinee j (0= Content 6, 1= content 4);

X_{ij9} is the content category of item i for examinee j (0= Content 6, 1= content 5);

X_{ij10} is the position of the correct answer of item i for examinee j (0 = D, 1 = A);

X_{ij11} is the position of the correct answer of item i for examinee j (0 = D, 1 = B);

X_{ij12} is the position of the correct answer of item i for examinee j (0 = D, 1 = C);

X_{ij13} is the figure status of item i for examinee j (0 = no figure, 1 = figure);

R_{ij} represents random error;

and,

β_{jp} are regression coefficients that characterize the structural relationship within person j;

[1] Relative item difficulty is calculated as final estimate of examinee ability minus item difficulty. Two additional variables were considered for the analysis—examinee ability and item difficulty. Due to the targeting of computerized adaptive tests, all three variables are highly correlated.

[2] Item length was calculated as the number of characters in the item including stem and all distractors.

for

$i = 1 \ldots n_j$ items taken by examinee j;

$j = 1 \ldots K$ examinees; and

$p = 0 \ldots P\text{-}1$ independent variables

The within person regression coefficients, β_{ij} are allowed to vary across examinees. For each of the P regression parameters in equation (1) we pose the following between-person model:

$$\beta_{jp} = \theta_{0p} + \theta_{1p}Z_{j1} + \theta_{2p}Z_{j2} + \theta_{3p}Z_{j3} + \theta_{4p}Z_{j4} + \theta_{5p}Z_{j5} + \theta_{6p}Z_{j6} + \theta_{7p}Z_{j7} + \theta_{8p}Z_{j8} + U_{jp} \qquad (2)$$

where

Z_{j1} is the test anxiety score for examinee j[3];

Z_{j2} is the ethnic background of examinee j[4] (0 = White, 1 = Asian);

Z_{j3} is the ethnic background of examinee j (0 = White, 1 = Other);

Z_{j4} is the gender of examinee j (0 = male, 1 = female);

Z_{j5} is the language status of examinee j (0 = English is the examinee's first language 1 = English is not the examinee's first language);

Z_{j6} is the age of examinee j (0 = Age 20-29, 1 = Age 30-39);

Z_{j7} is the age of examinee j (0 = Age 20-29, 1 = Age > 40);

Z_{j8} is the final ability estimate of examinee j;

U_{jp} represents random error;

---

[3] After taking the computerized adaptive test, examinees were asked to respond to a questionnaire on the computer. The questionnaire included demographic items and a 12 item test anxiety survey. The anxiety survey was scored 0-12 with 12 indicating the highest level of test anxiety (See Appendix for survey text).

[4] Examinees chose White, Black, Hispanic, Asian, or other. Due to the small numbers of examinees in the Black, Hispanic and Other categories, the data was coded as 0 = White, 1 = Asian and 0 = White and 1 = Other.

and,

θ_{qp} are regression coefficients that capture the effects of person-level variables on the structural relationships, β_{jp};

for

$q = 0 \ldots Q\text{-}1$ independent variables in the second stage model

Results

We began the analysis with the simplest possible model (Model 1). We posed a within-person model in which the outcome t_{ij} (the log of the response time for each item i for person j) varies around a person mean:

$$t_{ij} = \mu_j + R_{ij} \tag{3}$$

In the between-person model, each persons's intercept, μ_j, is modeled as a function of the grand mean, λ, plus a random error, U_j:

$$\mu_j = \lambda + U_j \tag{4}$$

This analysis is equivalent to a one-way analysis of variance with random effects (Bryk, et.al., 1989). Table 1 shows the results of the variance estimates for the within-person model and between-person model. The variance within persons is more than 8 times greater than the variance between persons.

INSERT TABLE 1 HERE

We then analyzed the data with increasingly complex models, adding the independent variables into the within-person model. Coefficients were initially allowed to vary from person to person. When the regression coefficient of an effect did not vary substantially across persons, the variable was treated as a fixed effect.

Tables 2 and 3 show the variance estimates and the regression parameter estimates when all within-person variables are added to the model (Model 2). In this model, all independent variables were significant predictors of the log of response time. All variables were modeled as fixed effects except for "figure". This means that while relative item difficulty, getting the item correct, item sequence, item length, content category and position of the correct answer did significantly predict variance in response time, these effects did not vary across persons. The presence of a figure significantly increases response times and the magnitude of this effect varied across persons. Therefore "figure" was modeled as a random effect.

INSERT TABLES 2 AND 3 HERE

Consistent with our earlier findings, relative item difficulty, whether the item was answered correctly, and item sequence were negatively related to the log of response time.[5] The harder an item was for the person, the longer they pondered it. Examinees spent more time on items they got wrong than on items they got right. Increased total text of the item resulted in increased time spent on the item.

Also consistent with our earlier findings, some content areas take significantly longer than others. Examinees take longer to respond when the correct answer is A, B, or C than when the correct answer is D.

A comparison of the within-person error variance estimates for Model 1 (.44) and Model 2 (.30) shows a significant reduction in the error variance as a result of the addition of the

[5] Since relative item difficulty is calculated at the person ability minus the item difficultly, a negative value indicates that the item was more difficult than the person was able.

within-person variables (See Tables 1 and 2).

To explore the range of response time variance explained by the within-person variables we ran 204 separate regression analyses. The mean R^2 was .36 (S.D. = .09) with a range of .11 to .59. For the fifty percent of examinees falling within the 1st and 3rd quantiles, the range was .31 to .43. Thus for most examinees the within-person variables explained at least one third of the variation in log response time.

We next added the between-person variables into the model. Only one variable, anxiety, significantly predicted variance between examinees. Accordingly, the final between-person model can be written as:

$$\beta_{jp} = \theta_{0p} + \theta_{1p} Z_{j1} + U_{jp} \tag{5}$$

where the within-person regression coefficients are expressed as a function of the mean of log response time (intercept), test anxiety, and random error. Table 4 shows the variance estimates for the final model (Model 3) while Table 5 shows the regression parameter estimates. Examinees who are more anxious take longer on their test. There was no interaction between figure and anxiety.

INSERT TABLES 4 AND 5 HERE

The likelihood ratio Chi-Square for Models 1, 2, and 3 were 32139.64 (d.f. = 203), 26278.37 (d.f = 190) and 26271.04 (d.f. = 188) respectively, indicating that each successive model significantly predicted more of the variance in log response time than the previous model.

Discussion

The major findings of this study relative to item level variables were consistent with our hypotheses and our earlier research (Gershon et.al., 1993). Response time increased proportionately with increasing item text length and increasing relative item difficulty. Item sequence also was an important factor in that response time was greater for *earlier* items in the test. It appears that persons may be more cautious at the beginning of the test or they may "warm-up" and move more quickly as they proceed through the test.

When distractor D was keyed as the correct response, response time was shorter than when A,B, or C was the correct response. We assumed that examinees would take longer to respond to items when the correct answer was D. The increased response time may be due to the procedure used for computerized testing. When an examinee pressed the space bar or down arrow key to highlight a response, distractor A was highlighted first. Subsequent key presses moved the highlight to B, C and finally D. Examinees may have read through each response while pressing the space bar and then returned to the answer when it was A, B or C (Dodd, 1993).

The presence of a figure was a major predictor of increased response time. This may have been due to the separate illustration booklet rather than the inclusion of a graph or table in the item. The magnitude of the effect varied across examinees. Some examinees may have had more trouble than others finding figures in the booklet, or the difference may have been due to how they perceived figures. Comparisons of response times with different graphic presentations (illustration booklet versus on-screen presentation) are a topic for future research.

Computer adaptive tests administer individualized combinations of items to each examinee. Therefore, if a test developer wishes to exert some control over total testing time, the entire bank of items can be reviewed. Items with characteristics that make them especially time consuming, such as long text can be deleted from the bank or rewritten. The distribution of keyed answers within the bank could be balanced across item difficulty to prevent, for example, many easy items keyed with a correct answer of D. The bank could also be reviewed to determine the number of figures associated with items by difficulty level.

A second method for controlling total test time is to constrain item selection. In this test, the adaptive algorithm balanced the content of items administered. The test administration software could also constrain the number of figures administered to an examinee or balance the answer key. Additional constraints on item selection would need to be carefully explored since limiting available items can impact upon test efficiency.

Modeling these data using a hierarchical linear model improved our understanding of the variance in response time. We had the expectation that "people are different". We expected some people to be "fast test-takers" and others to be "plodders". While there were significant differences between examinees, we found much more variation within examinees. For some examinees, item characteristics predicted over half of the variation in response time. For other examinees, we still have little understanding of why they spent more time on one item than on another item. Perhaps, regardless of particular item characteristics (eg. the item is relatively easy or the text is short), when examinees are unsure of an answer they spend a lot of time thinking about the content of the item.

An important finding of this study was that age, sex, first language, and ethnicity did not predict variance in response time. On this test, persons from particular demographic groups responded to items on the computerized test with comparable times. Final estimated ability was also not a significant predictor of response time. This means that low able examinees did not take longer than high able examinees. The adaptive nature of the test may account for this result since all examinees were challenged with items targeted to their ability level.

Test anxiety was the only between-person variable that was found to predict differences in response time. Persons who were more test anxious took longer to answer an item. Knowledge that test anxious examinees take more time may be useful to test developers.

Ultimately, a priori knowledge of response times may better enable test developers to predict the amount of time required for test administration based on fact rather than conjecture or tradition. Understanding how item characteristics impact on response time may allow test developers to allot total test time based upon the response time history of the individual test items.

Many testing organizations cite enhanced accessibility and improved security as reasons for administering tests by computer. While these are important factors in what is likely to be a universal move towards computerized testing in the future, having the capability to use advanced technologies to improve reliability and validity is another important goal. Understanding response time variance can help insure that "speededness" is not a factor in computerized adaptive achievement tests. This study also suggests that examinee characteristics are generally not related to response time, but that more controllable factors such as item length, position of the keyed correct answer, and use of figures do contribute to response time.

REFERENCES

Bryk, A.S., Raudenbush, S.W., Seltzer, M., Congdon, R.T., Jr., (1989). *An introduction to HLM* [Computer program and user's guide] Chicago, IL: Scientific Software, Inc.

Dodd, B. (1993) University of Texas. Personal communication.

Evans, F. and Reilly, R. (1976). *The LSAT speededness study-revisited.* Report #LSAC-72-3. In Law School Admissions Council, Reports of LSAC Sponsored Research, (2), 1970-74. Princeton, NJ: Law School Admissions Council.

Fair Test. (Fall, 1992) *Computerized testing: more questions than answers.* Cambridge, MA: National Center for Fair and Open Testing.

Gershon, Richard G. (1990). *CAT ADMINISTRATOR* [Computer Program] Chicago, IL.: Computer Adaptive Technologies, Inc.

Gershon, R., Bergstrom, B., and Lunz, M.E. (1993). *Computer adaptive testing: exploring examinee response time.* Paper presented at the American Educational Research Association Meeting, Atlanta, GA.

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

Lunz, M., Bergstrom, B. and Wright, B. (1992) The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement*(16), 1, 33-40.

Rafaeli, S. and Tractinsky, N. (1991). Time in computerized tests: a multi-trait, multi-method investigation of general knowledge and mathematical-reasoning on-line examinations. *Computers in Human Behavior.* (7), 215-225.

Thum, Y.M. (1994). *Analyzing individual variation in psychology: A multivariate hierarchical linear model for behavioral data.* University of Chicago: Unpublished doctoral dissertation.

Wild, C., Durso, R. and Rubin, D. Effect of increased test-taking time on test scores by ethnic group, years out of school, and sex. *Journal of Educational Measurement.* (19) 1, pp. 19-28.

Wright, B. D. and Stone, M. (1979) *Best Test Design.* Chicago, IL.: Mesa Press.

Acknowledgments

The authors wish to thank Yeow Meng Thum, University of Chicago, for running the HLM computations for the data analysis and for his insightful comments on earlier versions of this manuscript.

Author's Address

Send requests for reprints or further information to Betty Bergstrom,Ph.D., Director of Psychometric Services, Computer Adaptive Technologies, Inc., 2609 W. Lunt Ave, Chicago, IL 60614.

Figure 1

Distribution of Response Time and Log Response Time

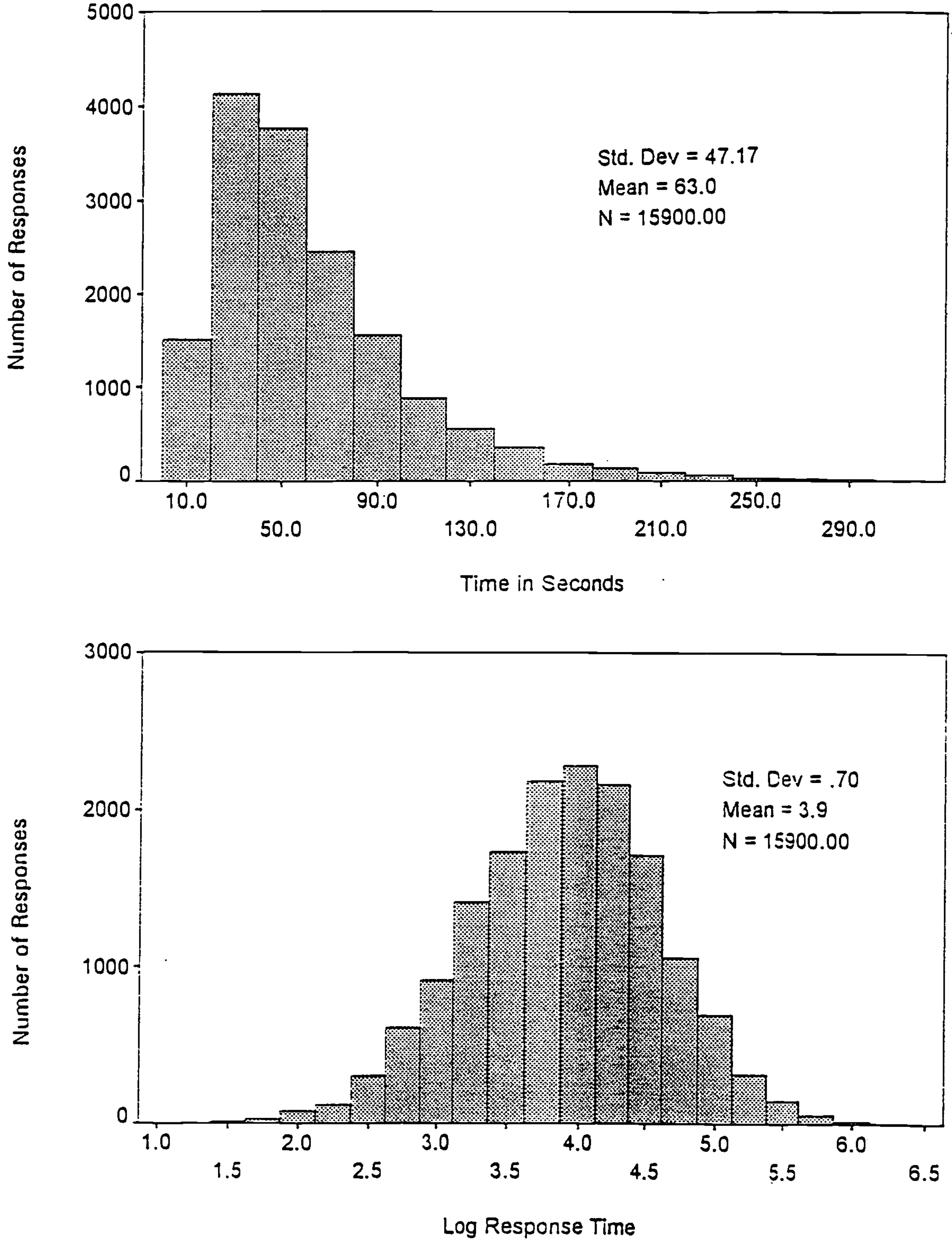

TABLE 1

Comparison of Person Variance Estimates for Model 1

	Within-Person Variance	
	Variance Estimate	Standard Error
Measurement Error	0.4414	(0.0050)

	Between-Person Variance	
	Variance Estimate	Standard Error
Mean Log Response Time (Intercept)	0.0454	(0.0051)

TABLE 2

Comparison of Person Variance Estimates for Model 2*

Within-Person Variance		
	Variance Estimate	Standard Error
Measurement Error	0.3017	(0.0035)

Between-Person Variance		
	Variance Estimate	Standard Error
Mean Log Response Time (Intercept)	0.0520	(0.006)
Mean/Figure	-0.0170	(0.004)
Figure	0.0101	(0.004)

*Within-person independent variables added to the model

TABLE 3

Regression Parameter Estimates for Model 2

		Estimate	S. Error	t-value	p-value
	RELATIVE ITEM DIFFICULTY	-0.0160	0.0025	-6.4000	0.0000
	CORRECT	-0.1917	0.0090	-21.3120	0.0000
	ITEM SEQUENCE	-0.0037	0.0002	-21.0489	0.0000
	ITEM LENGTH	0.0017	0.0000	59.6266	0.0000
	SUBTEST 1	0.2795	0.0140	19.9560	0.0000
Fixed	SUBTEST 2	0.0438	0.0174	2.5177	0.0118
	SUBTEST 3	0.1966	0.0141	13.8980	0.0000
	SUBTEST 4	0.1819	0.0144	12.6609	0.0000
	SUBTEST 5	0.1098	0.0173	6.3314	0.0000
	KEY POSITION A	0.0584	0.0130	4.4865	0.0000
	KEY POSITION B	0.0893	0.0126	7.0720	0.0000
	KEY POSITION C	0.0571	0.0123	4.6529	0.0000
Random	MEAN LOG RESPONSE TIME	3.2933	0.0252	130.7762	0.0000
	FIGURE	0.5716	0.0141	40.5598	0.0000

Degrees of freedom : 190

TABLE 4

Comparison of Person Variance Estimates for Model 3*

Within-Person Variance

	Variance Estimate	Standard Error
Measurement Error	0.3017	(0.0035)

Between-Person Variance

	Variance Estimate	Standard Error
Mean Log Response Time (Intercept)	0.0500	(0.006)
Mean/Figure	-0.0164	(0.004)
Figure	0.0099	(0.004)

*Within-person and between-person independent variables added to the model.

TABLE 5

Regression Parameter Estimates for Model 3*

		Estimate	S. Error	t-value	p-value
Fixed	RELATIVE ITEM DIFFICULTY	-0.0159	0.0025	-6.3729	0.0000
	CORRECT	-0.1915	0.0090	-21.2910	0.0000
	ITEM SEQUENCE	-0.0037	0.0002	-21.1037	0.0000
	ITEM LENGTH	0.0017	0.0000	59.6428	0.0000
	SUBTEST 1	0.2795	0.0140	19.9571	0.0000
	SUBTEST 2	0.0438	0.0174	2.5174	0.0118
	SUBTEST 3	0.1967	0.0141	13.9037	0.0000
	SUBTEST 4	0.1819	0.0144	12.6587	0.0000
	SUBTEST 5	0.1099	0.0173	6.3379	0.0000
	KEY POSITION A	0.0585	0.0130	4.4925	0.0000
	KEY POSITION B	0.0893	0.0126	7.0730	0.0000
	KEY POSITION C	0.0572	0.0123	4.6581	0.0000
Random	MEAN LOG RESPONSE TIME	3.1729	0.0509	62.3778	0.0000
	ANXIETY	0.0180	0.0066	2.7201	0.0065
	FIGURE	0.6105	0.0392	15.5731	0.0000
	FIGURE BY ANXIETY	-0.0058	0.0054	-1.0638	0.2874

Degrees of freedom : 188

*Anxiety was the only significant between-person independent variable.

APPENDIX

ANXIETY SURVEY

Response options: A. Yes B. No

1. If I were to take an intelligence test, I would worry a great deal before taking it.
2. During tests I find myself thinking of tasks unrelated to the task at hand.
3. I have an uneasy, upset feeling before taking an important test.
4. When taking a test, my emotional feelings interfere with my performance.
5. I seem to defeat myself while working on important tests.
6. As soon as a test is over I try to stop worrying about it, but I just can't.
7. Thinking about the score I may get, interferes with my studying and performance on tests.
8. If I knew I was going to take an intelligence test, I would feel confident and relaxed beforehand.
9. Even when I'm well prepared for a test, I feel anxious about it.
10. I dread courses where the professor has the habit of giving "pop" quizzes.
11. Thoughts of doing poorly interfere with my performance on tests.
12. I freeze up on things like intelligence tests and other important tests.

TM025556

NCME Annual Meeting, April 9-11, 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Computerized Adaptive Testing Exploring Examinee Response Time Using Hierarchical Linear Modeling

Author(s): Betty Bergstrom, Richard Gershon, Mary Lunz

Corporate Source:	Publication Date:
Computer Adaptive Technologies, Inc.	April, 1994

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here

Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here

Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: B. Bergstrom	Position: Dir. of Psychometric Services
Printed Name: Betty Bergstrom	Organization: Computer Adaptive Technologies
Address: 2609 W Lunt Ave #2E Chicago IL 60645	Telephone Number: (312) 274-3286
	Date: 4/22/96

OVER