DOCUMENT RESUME

ED 370 961 TM 021 500

AUTHOR	Parshall, Cynthia G.; And Others
TITLE	Response Latency: An Investigation into Determinants of Item-Level Timing.
PUB DATE	Apr 94
NOTE	20p.; 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/PC01 Plus Postage.
DESCRIPTORS	Ability; Adaptive Testing; Age Differences; Classification; College Students; Difficulty Level; Estimation (Mathematics); Ethnicity; Higher Education; *Item Response Theory; *Mathematics Tests; Probability; *Reaction Time; Responses; Sex Differences; *Test Items; *Timed Tests; Time Factors (Learning)
IDENTIFIERS	Accuracy; Response Patterns

ABSTRACT

Response latency information has been used in the past to provide information for consideration along with response accuracy when obtaining trait level estimates, and more recently, to flag unusual response patterns, to establish appropriate time-to-test limits (Reese, 1993), and to determine predictors of the amount of time needed to administer a given item (Gershon, Bergstrom, & Lunz, 1993). Data for this research were obtained from administration of an adaptive college mathematics placement test to 3,364 examinees. This study investigated item and examinee variables as potential influences of item-level response time. The item variables in this study were presentation order in the test, content classification, and cognitive classification. Examinee variables consisted of estimated ability, average rate of response, gender, ethnicity, age. The examinee-item variables were conditional probability and correctness of response. Results were analyzed through a series of regression models to ascertain the variables that function as the strongest determinants of item latency. There are six tables and three figures. (Contains 14 references.) (Author/SLD)

Response Latency: An Investigation into Determinants of Item-Level Timing

Cynthia G. Parshall
David J. Mittelholtz
Timothy R. Miller
American College Testing Program

Paper presented at the Annual Meeting of the National Council on Measurement in Education, April 4-8, 1994, New Orleans, LA.

Abstract

Response latency information has been used in the past to provide information for consideration along with response accuracy when obtaining trait level estimates (Tatsuoka & Tatsuoka, 1980; Thissen, 1983). More recently, interest has developed in using response latency for additional purposes. These uses include flagging unusual response patterns (Kingsbury, Zara, & Houser, 1993), establishing appropriate time-to-test limits (Reese, 1993), and determining predictors of the amount of time needed to administer a given item (Gershon, Bergstrom, & Lunz, 1993).

Data for this research were obtained from the administration of an adaptive college mathematics placement test. This study investigated item and examinee variables as potential influences of item-level response time. The item variables in this study were presentation order in the test, content classification, and cognitive classification. Examinee variables consisted of estimated ability, average rate of response, gender, ethnicity, age. The examinee-item variables were conditional probability and correctness of response. Results were analyzed through a series of regression models to ascertain the variables which function as the strongest determinants of item latency.

Response Latency: An Investigation into Determinants of Item-Level Timing

There has long been an interest in examining the amount of time a person takes to respond to a given stimulus. Research on this issue has proceeded in several directions. For example, early work utilized reaction time to investigate the cognitive processes and components involved in intelligent performance (e.g., Hunt, 1980; Sternberg, 1980). Hunt, Lunneborg, and Lewis (1975) investigated lexical access time (the time needed to assign meaning to a stimulus) using speed of response for stimuli of differing complexity, and found consistent correlations of around -0.30 between lexical access time and verbal intelligence. Sternberg's componential analysis approach (1980; 1984) uses latencies to investigate hypotheses about the components and processes involved in various types of reasoning, such as solving verbal analogies.

Another avenue of research involving response latency data concerns the effects of speededness on standardized tests and the relationship between timing effects and ethnic group membership (e.g., Llabre & Froman, 1987; Miller, Davey, and McLarty, 1992). Response latency data has also been used to enhance the information provided by response accuracy when obtaining trait level estimates (Tatsuoka & Tatsuoka, 1980; Thissen, 1983).

With the increasing number of computer-based test applications, and the computer's ability to collect precise latencies for each item administered, the educational community has become interested in using response latency for additional purposes. These uses include flagging unusual response patterns (Kingsbury, Zara, & Houser, 1993), establishing appropriate time-to-test limits (Bhola, Plake, & Roos, 1993; Reese, 1993), and determining predictors of the amount of time needed to administer a given item (Gershon, Bergstrom, & Lunz, 1993). With greater information about the functioning of response latency, latency information may be useful in identifying possible cheaters, or in determining non-motivated examinees in pre-test conditions. Latency data might also be useful for improving item selection and test construction procedures.

Although latency data can be collected under any computerized test administration, a computerized adaptive test will include a number of features that may impact the functioning of response latency. For example, under an adaptive administration different examinees are administered differing numbers of items. This clearly affects total test time, but after a given number of items, could affect item-level response time as well. In addition, under adaptive testing more proficient examinees will typically receive a smaller set of more difficult items (while less proficient examinees will receive a smaller set of less difficult items). Since more difficult items often require more time than easier items, response latency could vary across proficiency. Whether a given computer adaptive test permits or prohibits review of items may also have an effect on response latency. Item latency may also be differentially affected by whether a test is administered under timed or untimed conditions.

Purpose

This study was a preliminary investigation into the relationship between response latency and a number of examinee and item variables. Information about response latency determinants could be used to provide test administrators with the average predicted testing times needed for examination scheduling. For timed test conditions, this information could also be used as a baseline for determining an appropriate time length for the examination. Enhanced understanding about the relationships between response latency and examinee and item characteristics may also lead to improved test construction and perhaps to more accurate and efficient assessment.

Methods / Data Source

Data for this research were obtained from the COMPASS mathematics placement test, administered during the academic year 1992/93 at 28 colleges and universities across the country. The test was administered operationally as an adaptive test; the examinees in this study were administered different numbers of items, different orderings of items, and sometimes different mathematics domains.

The COMPASS mathematics test contains five domains: pre-algebra, algebra, college algebra, geometry, and trigonometry. Each domain has a pool of about 200 multiple-choice items. The test is adaptive at the domain level, as well as at the item level within domains. That is, examinees are adaptively (but transparently) routed through these predetermined domains, based upon their performance in earlier domains. Data from the algebra domain were used in the analyses. The item pool for this domain consists of approximately 250 items. Analyses were conducted on the 40 items administered most frequently to the examinees in this study. Examinees were each administered between 5 and 15 items in the algebra domain; however, a given examinee may also have been administered items from another mathematics domain, either before or after the algebra domain.

Testing information was available on 3,364 examinees. Although the intent of this study was to use operational data only, and the test itself was fully functional, some of the testing sites indicated that their administration purposes were exploratory. To the extent that examinees were aware of this, they may not have been motivated to do their best on the test. An initial preponderance of low test scores tended to support this hypothesis. In order to insure clean data for further analyses, certain heuristic were applied. If an examinee responded with any single option (e.g. 'A') 75% of the time or more, that score record was removed from the dataset; a total of 42 records were removed based on this rule. The responses of those examinees who took the entire algebra domain in less than 3 minutes and earned a percent-correct score of less than 40, were also removed (as evidencing a non-motivated form of test-taking). Sixty-two records were eliminated using this criterion. Problems with the test administration software necessitated the removal of 277 examinee records.

After all data-elimination heuristic had been applied, the mean algebra score for this sample was still quite low (on an estimated percent correct scale, the mean was 40 and the standard deviation was 19). However, further examination of the data failed to yield any additional clearly-defined, defensible rules for deleting cases. The final dataset consisted of 3,008 examinees. (Some records were flagged for deletion based on more than one of the heuristic listed).

This investigation was conceptualized as an exploratory study; thus, a variety of analyses were conducted to determine the relationships among variables. Several simple and multiple regression equations were computed to find those variables or variable combinations that were most strongly related to response latency. Significance testing was not considered especially important; given the large sample size of this study, the majority of analyses yielded significant results. Instead, an emphasis was placed on the meaningfulness and utility of results.

Results

The examinee variables investigated in this study were estimated ability (estimated theta), gender, ethnicity, age, and average rate of response. Each examinee's estimated theta was based solely on his or her performance on the algebra domain. The mean value of estimated theta for this sample was -0.26, with a standard deviation of 1.30. Males constituted 52% of the examinees, and females 48%. Ethnicity was a 9 option variable, including the option "prefer not to respond". Caucasians constituted the largest group (76.9%), followed by African Americans (7.5%) and Asian/ Pacific Islander (4.7%). The remaining groups each represented less than 2% of the sample. Although the majority of examinees were between 18 and 23 years of age, a substantial minority of the examinees (31%) were 24 or older. The final examinee variable investigated was average response latency; this average was computed across all items taken by an examinee, including those in other mathematics domains.

The item variables under investigation included presentation order in the test, content classification, and cognitive classification. Order ranged from position 1 to position 30, since some examinees were administered items in the algebra domain after they had responded to items in other mathematics domains. Content classification consisted of 12 levels, under the larger groupings of Coordinate Geometry, Elementary Algebra, and Intermediate Algebra. All items in the mathematics item pool are also classified according to cognitive level; the 3 levels of cognitive classification are Basic Skills, Application, and Analysis. Basic Skills items can be solved by performing a sequence of basic operations. Application items involve applying sequences of basic operations to novel settings or in complex ways. Analysis items require examinees to exhibit a conceptual understanding of the principles and relationships relevant to particular mathematical operations. All of the items analyzed in this study fell into two of those levels; 64% of the items were Basic Skills, and 36% were Application.

Two remaining variables were conceptualized as resulting from the interaction of examinee and item characteristics. One of these variables was the conditional probability of a correct response (computed on the examinee's estimated theta, and the item's IRT parameters). The final variable under analysis was the correctness of an item response, by a given examinee. As would be expected with an adaptively administered examination, approximately 50% of the items fell into each level of this variable.

The means and standard deviations of response latency, overall and by categorical variable, are provided in Table 1. The first column reports Ns. For the variables in the top half of the table this column refers to number of examinees; for those in the bottom half it refers to number of responses. The next two columns of this table display the mean and standard deviation of the raw latency values in seconds (e.g., males had an average response latency across all items of 81.68 seconds, with a standard deviation of 42.03 seconds). The last two columns display the mean and standard deviation of a transformation of the latency values. The latency data in this study were highly skewed, as is often the case with latency data. A log transformation yielded a distribution which was highly peaked, but no longer skewed. This transformation was applied to both the average latency for an examinee across his or her set of test items, and the individual latency for each specific item by each examinee. The log of the item latency and the log of an examinee's average latency were then used in all further analyses. The large standard deviations noted for both raw and transformed latency values suggest that response latency is highly variable.

The correlations among continuous variables are presented in Table 2. Most of these simple correlations are modest at best. The low correlations between predictor variables indicates that multicollinearity will not be a problem for the regression analyses. However, the low correlations between predictors and latency suggest that a powerful model for predicting latency from these variables may not be present.

Results were analyzed through regression models to ascertain which, if any, of the variables were strong determinants of item latency. The first regression analysis was a block, stepwise procedure. Variables were defined as blocks (based on their operational definition as Examinee, Item, or Examinee-Item characteristics) which were then entered into the model in these sets. Results of this regression equation are given in Table 3. The first block to enter the equation is the Examinee block, which yields an R^2 value of .26; when the Item and Examinee-Item blocks are added to the model, the R^2 value of the full model is .35.

In order to better interpret the results of the block, stepwise regression, a standard multiple regression with all variables in the model was then conducted. The R^2 for this model is .35 (see Table 4), indicating that a considerable proportion of variance in item response latency is left unaccounted for by the model.

The simple correlations provided in Table 2 suggest that average latency functions as the strongest predictor in this multiple regression. In order to analyze the strength of the remaining variables for explaining latency, a second multiple regression was conducted. This regression model included all predictor variables except average latency; the results are given in Table 5. The R^2 for this model is .14; considerably less variance in response latency is explained with average latency excluded from the model.

A more detailed analysis of the relationships between predictor variables and latency is provided in Table 6. The simple r^2s provided in this table confirm that the strongest relationship investigated is between average latency and latency. A positive correlation (r=.50) is found between latency and average latency, and the r^2 for the simple regression is .256. There is some dependency between these two variables, but this result also suggests that, of the variables examined in this study, the best explanatory variable of an examinee's rate of response to a given item may be the examinee's average rate of response.

Further examination of the results in Table 6 suggests three levels of relationship, with average latency representing the strongest level. Low but discernible relationships can also be found between latency and estimated theta, content classification, order, and conditional probability. The positive correlation between latency and the examinee variable of estimated theta (r = 0.21), suggests that as proficiency increases, so does response time. This relationship may be related to the fact that in an adaptive test, more-able examinees are typically presented with more difficult items (or, fewer overly easy items). A second possibility is that non-motivated examinees could have taken the test inattentively and rapidly, resulting in low ability estimates. Both of these occurrences would increase the relationship displayed between estimated theta and latency. The two item variables which show a modest relationship with latency are content classification and order. Mean latencies by algebra content classifications can be found in Table 1. Generally, these mean latencies increase with content complexity. However, a fuller examination of the relationship between content classification and latency would necessitate an item-by-item evaluation of the number of steps typically required to solve the problem. This can vary considerably within content classification, (i.e., it is possible to write very easy or very difficult items involving exponents). The negative correlation between latency and order indicates that examinees tend to take more time to respond to an item early in testing, and to respond more rapidly as testing continues. Detailed examination of item latency means reveals that the first two items have much larger mean latencies than later items, suggesting that one or two items may be needed for examinees to develop a familiarity with the computerized mode of responding. A second, smaller decrease in latency occurs around item 24, for those examinees who were administered more than one domain. Under these untimed conditions, these appeared to be the strongest item order effects. The examinee-item variable, conditional probability, also displays a modest relationship with latency; the correlation between these two variables is 0.11.

The remaining variables in the model show no meaningful relationship with latency in this sample of examinees. These variables are: ethnicity, age, gender, cognitive classification, and correctness of response. The very low contribution of examinee demographic variables to the model supports results found by Gershon, et al. (1993), who found that when demographics were added to a regression model which already contained item variables, they added less than 2% to the variance explained by the model. Although the mean differences on latency between ethnic groups appear large enough to be meaningful (see Table 1), a fuller examination of the relationship between latency, ethnicity, and proficiency reveals an interaction effect. Figure 1 reveals that for some ethnic groups, a higher response latency mean was associated with a lower average test score; for other groups a greater response latency mean was associated with a higher test score mean. The mean difference in latency between levels of the item variable, cognitive classification, also appears both meaningful and interpretable. The correlation ($r = -0.17$) between cognitive classification and latency indicates that items in the Application level require more time than those in the Basic Skills level. This is in keeping with the rules used to classify items into these two cognitive categories. The major difference between the Basic Skills and Application cognitive levels is that Application items involve more steps or more novel settings, both of which might be expected to require more time for solution. Apparently, the large standard deviation mediates this effect, resulting in a low r^2 value.

A very low effect ($r^2 = .005$) is also displayed by the examinee-item variable correctness of response as a determinant of response latency. Previous research has been inconclusive regarding the relationship between correctness of response and latency. Kingsbury, et al. (1993) found correctness or incorrectness of examinee responses to be an important predictor of response latency. However, Gershon, et al. (1993) included correctness of response among examinee variables in their model, and found very little effect for that variable set. As a follow-up analysis in this study, the mean latencies for correct versus incorrect responses were computed for examinees in three categories of total algebra score. These categories were defined as the upper and lower 25% of examinees, and the middle 50% of examinees. Figure 2 reveals that examinees in the upper 25% scoring group had a sizeable mean difference on latency between items which they answered correctly as opposed to those they answered incorrectly. Response latencies for these examinees were considerably longer for incorrect items than for correct items. This mean difference is smaller for the middle 50% scoring group; for the lower 25% scoring group the mean difference is very small, and in the opposite direction.

Conclusions

Response latency appears to be a highly variable trait, both across items and across examinees. The explanatory variables in this study did not yield a strong model for predicting response latency, either individually or in combination. Although the potential for future applications of latency data may be promising, the level of information currently available does not lend itself to easy utilization.

It is possible that stronger relationships may be found under other conditions; various aspects of the test administration could have affected latencies. Computer-based testing was almost certainly a new experience for the majority of examinees. When the testing mode is not such a novel experience, a clearer latency effect may be found. Certainly, knowledge that the test was unspeeded could affect examinees' self-pacing during testing (Llabre & Froman, 1987), and this would relate directly to the response latencies collected. An interesting avenue for future research might be the differing latency results obtained under test conditions where item review is allowed as or osed to those testing conditions where an examinee is only given a single opportunity to respond to each item. These test design features may well interact with the test-taking strategies an examinee utilizes (Parshall & Kromrey, 1993), which will then impact response latency.

Another direction for future research suggested by this study is the interaction between ethnic group and latency. It may be that the score differences found often found between various demographic groups were mediated in this study by the untimed conditions (Buhr & Legg, 1989). The minimal regression effect for examinee demographic characteristics and latency was also found in an earlier study (Gershon, et al., 1993), and offers encouraging results regarding equity issues and computer testing. An important follow-up study would be one in which latency, conditioned on ability, is compared across ethnic group on an item-by-item basis (i.e., members of the two groups would be matched on ability first, and then latency would be compared).

References

Buhr, D. C. & Legg, S. M. (1989). Development of an adaptive test version of the College Level Academic Skills Test. Contract #88012704. Institute for Student Assessment and Evaluation, University of Florida.

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 Educational Research Association, Chicago.

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.

Hunt, E. B. (1980). Intelligence as an information-processing concept. British Journal of Psychology, 71, 449-474.

Hunt, E. B., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? Cognitive Psychology, 7, 194-227.

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.

Llabre, M. M. & Froman, T. W. (1987). Allocation of time to test items: A study of ethnic differences. Journal of Experimental Education, 55, 137-140.

Miller, S., Davey, T., & McLarty, J. (1992, April). Times test performance of black and white examinees: An investigation of differential response strategies. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Parshall, C. G. & Kromrey, J. D. (1993). Computer testing vs. paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

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

Sternberg, R. J. (1980). Sketch of a componential sub-theory of human intelligence. Behavioral and Brain Sciences, 3, 573-614.

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. Behavioral and Brain Sciences, 7, 269-287.

Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing. New York: Academic Press.

Tatsuoka, K. K. & Tatsuoka, M. M. (1980). A model for including time data in scoring achievement tests. In D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis. MN: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Table 1
Means and Standard Deviations of Latency and Log-Latency Overall and by Groups

		Latency		Latency (Log)	
	N	Mean	Standard Deviation	Mean	Standard Deviation
Overall	3008[a]	79.34	70.46	4.06	0.83
Gender					
Male	1560	81.68	42.03	4.08	0.52
Female	1447	80.64	38.16	4.10	0.47
Ethnicity					
Black African-American	225	86.05	43.17	4.14	0.53
American Indian-Alaska Native	34	81.59	44.83	4.09	0.55
White/Caucasian	2314	78.70	39.34	4.06	0.49
Mexican-American/Chicano	57	88.11	38.28	4.10	0.61
Asian/Pacific Islander	142	97.54	39.99	4.29	0.46
Puerto Rican/Cuban/Other Hispanic	44	89.34	40.14	4.20	0.44
Filipino	33	106.56	58.80	4.32	0.49
Other	45	92.71	39.40	4.22	0.48
Prefer Not to Respond	114	83.58	37.37	4.09	0.55
Correctness of Response					
Correct	15223[b]	79.46	67.06	4.12	0.73
Incorrect	15395	79.21	73.67	4.01	0.91
Cognitive Classification					
Application	11099	83.70	70.31	4.13	0.81
Basic Skills	19519	76.85	70.42	4.02	0.84
Content Classification					
Coordinate Geometry					
Distance formula in the plane	1791	86.15	70.91	4.20	0.73
Linear equations in 2 variables	1975	58.27	47.32	3.79	0.78
Elementary Algebra					
Basic operations with polynomials (+,-,x,/)	1102	74.19	55.46	3.99	0.90
Exponents	3309	69.51	62.81	3.92	0.85
Formula manipulation and field axioms	895	60.25	39.04	3.92	0.61
Factorization of polynomials	2623	116.70	89.52	4.49	0.79
Linear inequalities in 1 variable	5337	77.15	60.51	4.07	0.79
Solution of polynomial equations by factoring	1824	115.82	103.56	4.46	0.77
Setting up equations for given situations	2199	60.55	45.24	3.91	0.64
Substitution of values into algebraic expressions	4987	73.04	57.61	3.99	0.85
Intermediate Algebra					
Exponents	2854	96.51	94.70	4.15	0.96
Rational expressions	1722	53.51	43.94	3.76	0.66

Note: [a] N refers to number of examinees
[b] N refers to number of responses

Table 2
Correlation Matrix

	Latency (Log)	Average Latency (Log)	Estimated Theta	Order	Conditional Probability	Age
Latency (Log)	1.0					
Average Latency (Log)	0.50	1.0				
Estimated Theta	0.21	0.35	1.0			
Order	-0.17	0.00	-0.09	1.0		
Conditional Probability	0.11	0.26	0.69	-0.10	1.0	
Age	0.06	0.17	-0.13	0.08	-0.08	1.0

Note: N = 30618; except for Age where N = 29402

Table 3
Stepwise Multiple Regression of Blocked Examinee, Item, and Examinee-Item Variables on Latency

Variable	Step	R^2	F	p
Examinee	1	.26	855.63	0.000
Item	2	.34	269.96	0.000
Examinee-Item	3	.35	226.63	0.000

Table 4
Multiple Regression of All Examinee and Item Variables on Latency

Model	$R^2 = .35$	F = 580.82	$p \leq 0.000$
Variable	**Parameter Estimates**	**t**	**p**
Intercept	0.12	2.48	0.01
Average Latency	0.95	95.79	0.00
Estimated Theta	0.09	18.48	0.00
Gender	-0.01	-1.78	0.08
Age	0.00	-2.14	0.03
Ethnicity1	10.06	2.59	0.01
Ethnicity2	0.09	2.17	0.03
Ethnicity3	0.02	1.18	0.24
Ethnicity4	-0.02	-0.67	0.50
Ethnicity5	0.04	1.32	0.19
Ethnicity6	0.01	0.25	0.80
Ethnicity7	-0.03	-0.65	0.51
Ethnicity8	0.04	1.12	0.26
Order	-0.02	-29.64	0.00
Cognitive Classification	0.19	16.57	0.00
Content Classification1	0.36	15.36	0.00
Content Classification2	0.27	11.92	0.00
Content Classification3	0.52	19.54	0.00
Content Classification4	0.23	11.07	0.00
Content Classification5	0.16	5.71	0.00
Content Classification6	0.73	33.62	0.00
Content Classification7	0.30	13.86	0.00
Content Classification8	0.67	29.26	0.00
Content Classification9	0.10	3.98	0.00
Content Classification10	0.49	24.08	0.00
Content Classification11	0.37	17.25	0.00
Correctness of Response	0.09	9.41	0.00
Conditional Probability	0.60	-21.29	0.00

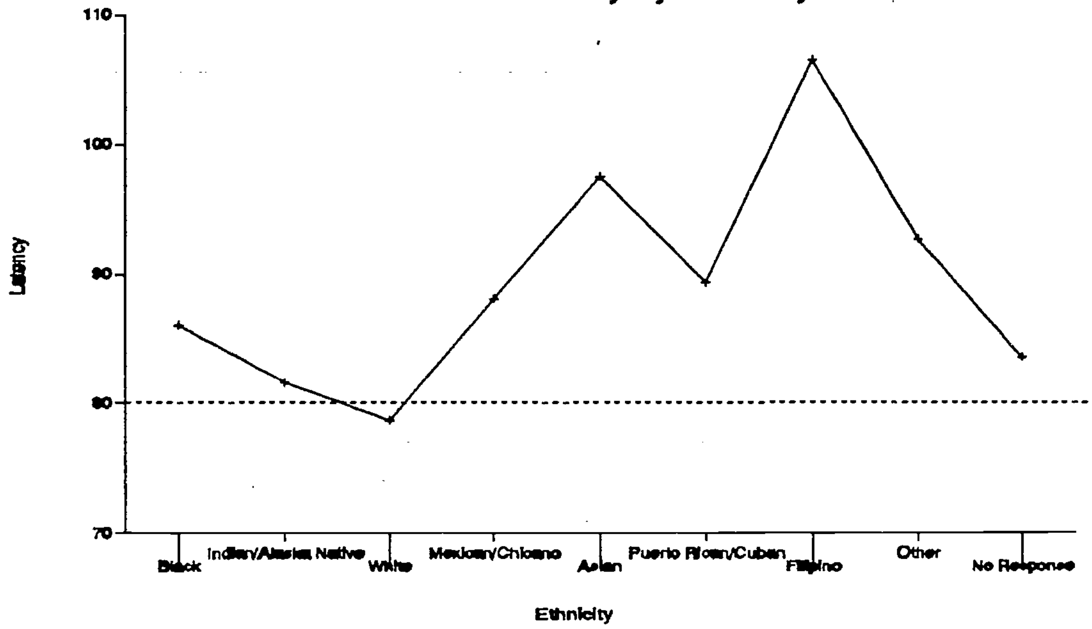
Table 5
Multiple Regression of Examinee and Item Variables (Except Average Latency) on Latency

Model	$R^2 = .14$	F = 190.69	$p \leq 0.000$
Variable	**Parameter Estimates**	**t**	**p**
Intercept	3.82	103.52	0.00
Estimated Theta	0.20	35.78	0.00
Gender	-0.02	-2.07	0.03
Age	0.01	18.51	0.00
Ethnicity1	0.13	4.72	0.00
Ethnicity2	0.05	1.01	0.31
Ethnicity3	0.02	0.75	0.46
Ethnicity4	0.03	0.85	0.40
Ethnicity5	0.19	5.94	0.00
Ethnicity6	0.21	4.81	0.00
Ethnicity7	0.21	4.32	0.00
Ethnicity8	0.07	1.64	0.10
Order	-0.02	-23.46	0.00
Cognitive Classification	0.21	15.40	0.00
Content Classification1	0.36	13.61	0.00
Content Classification2	0.24	9.00	0.00
Content Classification3	0.49	16.04	0.00
Content Classification4	0.22	9.10	0.00
Content Classification5	0.16	5.03	0.00
Content Classification6	0.74	29.96	0.00
Content Classification7	0.28	10.99	0.00
Content Classification8	0.66	24.92	0.00
Content Classification9	0.07	2.42	0.02
Content Classification10	0.44	18.95	0.00
Content Classification11	0.35	14.61	0.00
Correctness of Response	0.09	7.92	0.00
Conditional Probability	-0.51	-15.85	0.00

Table 6
Simple Correlations and Simple Regressions of Independent Variables on Latency

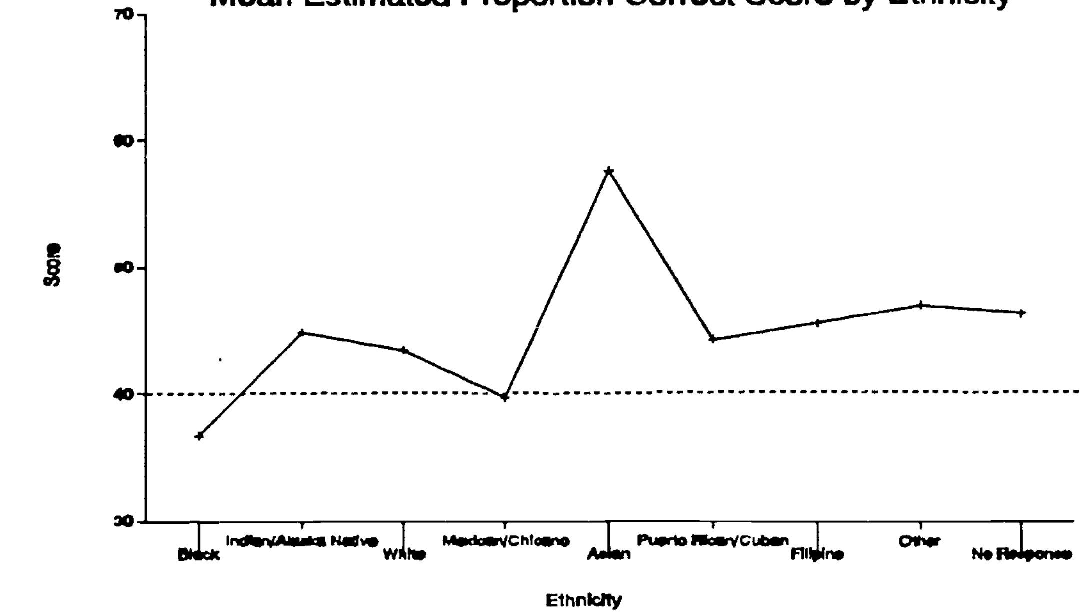
Independent Variable	Simple Correlation	Simple R-Square
Average Latency (Log)	0.50	0.256
Estimated Theta	0.21	0.043
Ethnicity	n/a	0.006
Age	0.06	0.004
Gender	n/a	0.000
Content Classification	n/a	0.063
Order	-0.17	0.028
Cognitive Classification	n/a	0.005
Conditional Probability	0.11	0.012
Correctness of Response	n/a	0.005

Mean Item Latency by Ethnicity

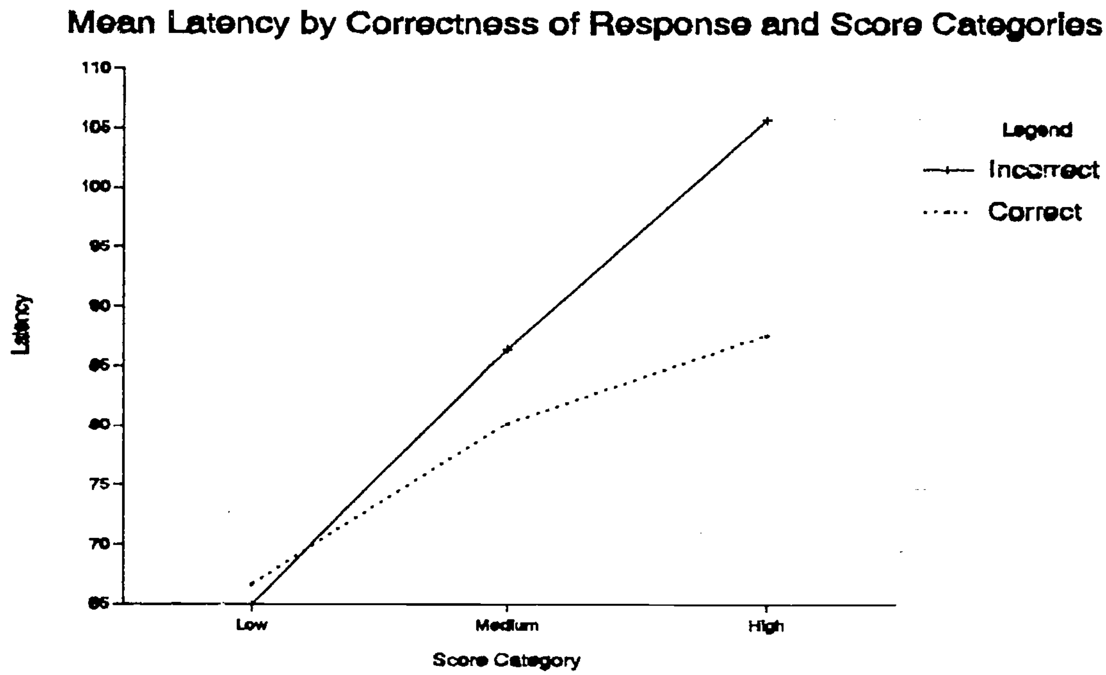


Note; ---- marks the overall mean latency

Mean Estimated Proportion Correct Score by Ethnicity



Note; ---- marks the overall mean score


Mean Latency by Correctness of Response and Score Categories
110
105
100
95
90
85
80
75
70
65
Latency
Low
Medium
High
Score Category
Legend
Incorrect
Correct