EFFECTS OF COMPUTERIZED ADAPTIVE TESTING ON BLACK AND WHITE STUDENTS

Steven M. Pine
Austin T. Church
Kathleen A. Gialluca
and
David J. Weiss

RESEARCH REPORT 79-2
MARCH 1979

PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

Prepared under contract No. N00014-76-C-0244, NR150-383
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Research Report 79-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Effects of Computerized Adaptive Testing on Black and White Students		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Steven M. Pine, Austin T. Church, Kathleen A. Gialluca, and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0244
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.:NR150-383
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE March 1979 13. NUMBER OF PAGES 47
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ability tests
adaptive tests
tailored tests
conventional tests
bias-reduced tests
psychological reactions
mode of administration
order of administration
knowledge of results
nervousness
motivation
guessing
bias
item bias
test bias
ICC theory
race

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Bias-reduced and non-bias-reduced conventional paper-and-pencil and computerized adaptive tests of word knowledge were administered to Black and White high school students to study differential effects on ability estimates and psychological reactions. Independent variables examined were bias-reduction, the presence or absence of knowledge of results after each item, mode of administration (paper-and-pencil or computerized adaptive), order of administration, and race. Dependent variables were three test performance variables (the ability estimates derived from both conventional paper-and-pencil and computerized adaptive tests, the variance of those estimates, and the number of omitted responses) and four psychological reaction variables (reaction to knowledge of results, nervousness, motivation, and guessing). Bias-reduced tests were specially constructed from items which had previously been shown to be less biased towards Black students in terms of an item bias index derived from item characteristic curve (ICC) theory. The bias-reduced tests eliminated mean racial differences between Black and White students under certain test conditions, but the effect interacted with other conditions of test administration, e.g., whether or not knowledge of results was provided. Since the bias-reduced tests provided less precise measurement than the non-bias-reduced tests, it was concluded that more traditional item statistics, such as item discriminations, should be considered along with an index of item bias in test construction. Computerized adaptive tests were generally shown to be more motivating than the conventional paper-and-pencil tests. Black students, in particular, seemed to be less tolerant of the conventional paper-and-pencil tests, especially when taken after the adaptive test. This was reflected in levels of reported motivation, number of omitted responses, and reported amounts of guessing. Differential psychological reactions for Black and White students were found for other conditions of test administration as well; however, the computer-administered adaptive tests appeared to reduce these differences in comparison to the conventional paper-and-pencil tests. These data imply the need for further study of the effects of test administration conditions on members of minority groups to determine those administration conditions which maximize ability estimates directly or through their effects on the psychological environment of testing.

CONTENTS

Acknowledgements

We would like to thank the Minneapolis Public Schools for their approval and support of this research. We are particularly grateful to the students and staff of Central and North High Schools for their enthusiastic cooperation. Special thanks are also due the Association of Afro-American Educators for their support.

Technical Editor: Barbara Leslie Camm

EFFECTS OF COMPUTERIZED ADAPTIVE TESTING ON BLACK AND WHITE STUDENTS

Because computerized adaptive or tailored testing has the capability of individualizing ability tests to the characteristics of an examinee, it would appear to have the potential for reducing group differences in test scores resulting from individual or group difference variables other than those that the test is designed to measure. These variables might include group differences in motivation, test-taking anxiety, or tendency to guess or to omit items.

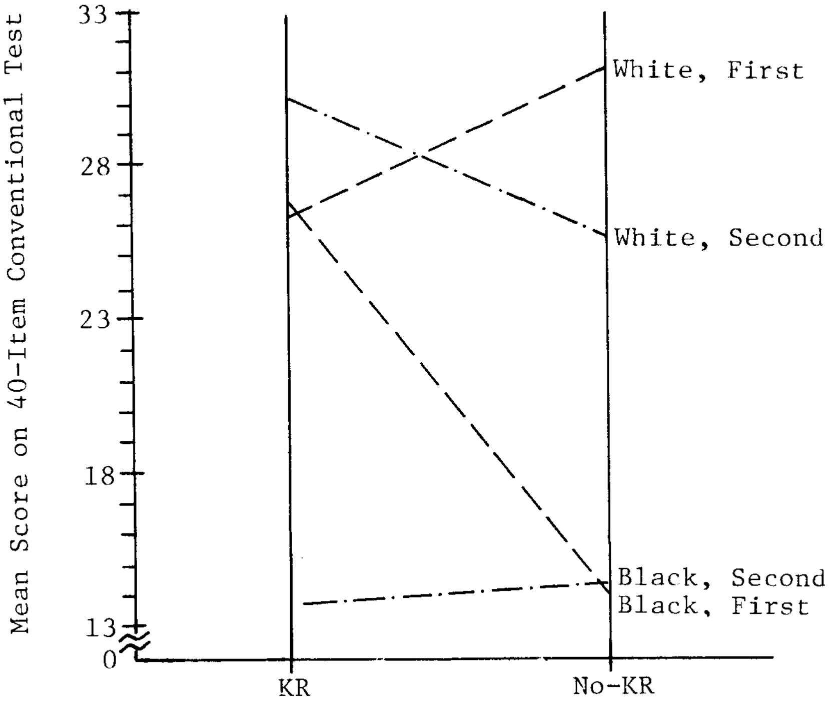
In conventional ability testing, items of the same difficulty are given to all examinees, regardless of their true ability levels. This reduces test reliability; consequently, the validity of the test may also be reduced in those groups which receive items inappropriate for their ability levels. Subgroups of the general population often differ with respect to background variables other than ability which may affect their performance on ability tests; therefore, test items which are appropriate in content for one subgroup may be inappropriate for another subgroup. With adaptive testing, it is possible to administer only those items that are appropriate for each group being tested. The process of adapting the test to each individual may also result in differential psychological impact on examinees from different population subgroups.

Previous research has provided some evidence for these potential psychometric and psychological benefits to minority examinees using computerized testing. Pine and Weiss (1978) demonstrated through a computer simulation that a Bayesian version of an adaptive test could reduce test unfairness within a simulated employee selection situation. In a live administration of computer-administered conventional tests, Johnson and Mihal (1973) administered identical conventional tests to Black and White students by paper and pencil and by computer. White students scored significantly higher than Black students on the paper-and-pencil tests, but not on the computer-administered tests.

In a study reported by Betz (1975, p. 24), two tests were administered by computer to a group of about 100 high school students, consisting of Black and White students. Both a conventional test and a pyramidal adaptive test (Larkin & Weiss, 1974) were administered to each student; half the group received the conventional test first, and half received the adaptive test first. In addition, half the group received feedback after each item indicating whether or not their answers were correct (knowledge of results, or KR, condition); the other half received no feedback after each test item (no knowledge of results, or No-KR, condition). The design was, therefore, a 2×2×2 analysis of variance. The independent variables were (1) race--Black and White, (2) knowledge of results (KR)--immediate or none, and (3) order--conventional test administered first or second. The data were analyzed for the conventional test only; thus, the dependent variable in this analysis was number-correct score on the conventional test.

The results for the three-way analysis of variance showed that the only significant main effect was for race. However, there was a significant three-way Order×Race×KR interaction. When a conventional test was administered first under conditions of immediate feedback, the mean of the Black students (26.4) was not significantly different from the mean of the White students (26.0), as is indicated in Figure 1.

Figure 1
Mean Scores for Black and White Students Completing a 40-Item Conventional Test First and Second in Both Knowledge of Results (KR) Conditions



If this result can be replicated, it implies that race differences observed in test scores may be a function, not of differences in ability levels, but of differences in the psychological effects of the conditions of administration. These findings, although not completely replicating those of Johnson and Mihal (1973), do support their general conclusion that conditions of test administration might affect motivational conditions, which in turn may reduce race group differences to nonsignificant levels.

The purpose of the present study was to replicate and to extend the previous findings that computerized administration of ability tests can increase the test scores and the test-taking motivation of minority examinees. Specifically, the present study compared a computerized adaptive test designed to minimize test bias with a similar conventional paper-and-pencil test in order to investigate possible racial differences on the following variables:

1. Test performance variables
   a. Ability test scores
   b. Standard errors of measurement
   c. Number of omitted responses

2. Psychological reaction variables
   a. Reaction to knowledge of results
   b. Test-taking anxiety (nervousness)
   c. Motivation
   d. Tendency to guess.

METHOD

Subjects

Two hundred and thirty-four students from a Minneapolis high school were tested. Black and White students were about equally represented in the total group. A small amount of subject attrition occurred because of equipment failures and interruptions unrelated to the testing procedure, thus resulting in incomplete data sets. The number of missing subjects differed for different analyses and therefore is reported separately for each analysis. Each student was tested during the course of a normal school day and received a McDonald's gift certificate worth $.50 for participating in the study.

Design

The design for this study was a five-way factorial with repeated measures on one factor; the other four variables were completely crossed. Table 1 summarizes the five independent variables. Each student was assigned sequentially to one of the bias-reduction (BR)×knowledge of results (KR)×

Table 1
Description of Independent Variables

Independent Variable	Number of Conditions	Conditions	Type of Variable
Bias-Reduction (BR)	2	Bias-Reduced, Non-Bias-Reduced	Crossed
Knowledge of Results (KR)	2	Immediate Knowledge of Results, No Knowledge of Results	Crossed
Mode of Administration	2	Computer-Administered, Paper-and-Pencil	Repeated
Order of Administration	2	Paper-and-Pencil Test First, Computer-Administered Test First	Crossed
Race	2	Black, White	Crossed

order conditions within his/her respective racial group. The student was then administered two vocabulary tests--one conventional paper-and-pencil test and one computerized adaptive test--in the appropriate order. The major dependent variable derived from these tests was the student's ability level estimate obtained by scoring procedures based on item characteristic curve (ICC) theory. The number of omitted responses in each test was also recorded for each student. In addition to the vocabulary tests, each student was administered a test reaction questionnaire after each test condition.

Independent Variables

Bias Reduction

Item pool. The item pool consisted of 187 five-alternative multiple-choice word knowledge items gathered from several sources. Seventy-six of these items were developed and/or parameterized by Church, Pine, and Weiss (1978). Of these 76 items, 32 were written specifically as "Black-type" words; that is, it was assumed that the Black students would have greater familiarity with them than would the White students. Similarly, an additional 17 items were chosen as "White-type" words. Examples of each of these item types are given in Appendix Table A. The items not taken from Church et al. (1978) were obtained from the University of Minnesota computerized adaptive testing vocabulary item pool (McBride & Weiss, 1974).

For each item, item calibration procedures (see Church et al., 1978, pp. 19-22) yielded an index of bias and two standard ICC parameters (discriminating power, a, and item difficulty, b). The third ICC parameter, c, was set to .20 for all items, which is equal to 1 divided by the number of response alternatives. Bias was indexed by an ICC version of the Angoff and Ford (1971) elliptical distance measure of item bias (Martin, Pine, & Weiss, 1978). Since the elliptical distance index is highly correlated with the difference between the ICC item difficulties of the two contrasted groups, bias was indexed in the present study by the difference between the item difficulty (b) values for the Black and White groups. A positive value of the bias index indicates an item biased against the minority group, while a negative value indicates an item biased against the majority group. The calibrated item pool was then used to form two conventional paper-and-pencil tests and two computer-administered adaptive tests.

Computerized adaptive tests. The computer-administered adaptive tests (CAT) were constructed using the stradaptive testing strategy (Weiss, 1973). All items were assigned to one of seven strata based on the difficulty (b) parameter. Appendix Table B gives the a and b parameters and bias index for each item in the stradaptive pool.

To begin the stradaptive test, an initial stratum assignment was made by asking the students to rate themselves on verbal ability on a 3-point scale. Each student was asked the following question:

> Compared to other people, how good do you think your vocabulary is?
> 1. better than average, 2. average, 3. below average.

He/she was told to type a number from "1" to "3" accordingly. Students were then given the first item in Stratum 6, 4, or 2, depending on their respective self-ratings. In accordance with usual stradaptive item selection procedures, students were subsequently administered items from the next-more-difficult or next-less-difficult stratum, depending on whether the response to the previous item was correct or incorrect. Each stradaptive test was terminated after 20 items.

Two forms of the adaptive test were constructed from the same item pool. In the bias-reduced (BR) adaptive test, items were arranged within each stratum in increasing order of bias. In the non-bias-reduced (NBR) adaptive test, items were arranged within each stratum in decreasing order of item discrimination, following recommendations for the construction of stradaptive tests (Weiss, 1974). Thus, in the BR condition, each item administered was the item with the lowest bias value still available in the appropriate stratum. In the NBR condition, each item administered was the most discriminating item remaining in the stratum.

Conventional paper-and-pencil tests. Two 20-item conventional paper-and-pencil (P&P) tests--one bias-reduced (BR) and one non-bias-reduced (NBR)--were constructed using items not used in the stradaptive test item pool. Item parameters and bias indices for these items are shown in Appendix Table C. The BR test included items with low positive or negative values of the bias index, while the NBR test included items with higher positive values of the bias index. Each set of 20 items formed a peaked test, with item difficulty peaked at the level of difficulty of Stratum 4, the middle stratum.

In order to equate conditions for the conventional paper-and-pencil and computer-administered adaptive tests as much as possible, items for the BR paper-and-pencil tests were selected to have approximately the same item bias values as the first few items that would be administered in each stratum of the BR adaptive test, and items for the NBR paper-and-pencil test were selected to have approximately the same item discrimination values as the first few items in each stratum of the NBR adaptive test. Consistent with this test-construction strategy, some items could be used in both the computerized tests and the paper-and-pencil tests as long as they were not in the same BR condition in both modes, since each student took the computerized and paper-and-pencil tests under only one BR condition.

It was impossible to match exactly the item characteristics of the 20 items in the conventional paper-and-pencil tests to the 20 items actually administered by the computerized adaptive tests, since it could not be determined in advance exactly which 20 items would be administered in the adaptive test to each student. Consequently, in order to compare these two testing strategies, the item characteristics of the computer-administered adaptive tests were calculated after administration of the tests (see Table 2 below).

Mode and Order of Administration

Each student completed a computer-administered test (adapted to his/her ability level) and a conventional paper-and-pencil test, both of which were either bias-reduced (BR) or non-bias-reduced (NBR). Half of the students took the paper-and-pencil test first (Order 1), while the other half took the computer-administered test first (Order 2).

The adaptive tests were computer administered by cathode-ray terminals (CRT) connected by telephone to a real-time computer system using procedures similar to those described by DeWitt and Weiss (1974). Each test item was presented separately on the CRT screen at the rate of 30 characters per second. Students were told that they could type a question mark in response to an item if they did not know the answer and wanted to omit it.

The paper-and-pencil tests were administered in booklets especially prepared for this study. Students had ample time to complete the tests and were instructed to omit an item if they did not know the correct answer.

Knowledge of Results

For half the students, immediate knowledge of results (KR) was administered after each test item, indicating whether or not the student's answer was correct; the other half received no information concerning the correctness of their answers (No-KR).

For the computer-administered tests, either the word *Correct* or *Incorrect* appeared on the screen after the student responded. The student then typed the letter *P* (for proceed) on the CRT keyboard in order to have the next question presented. In the No-KR condition, the next question appeared immediately after the student's answer was typed. KR in the paper-and-pencil mode was given using a latent ink process. Students marked their answer sheets with a special pen causing a latent image, which was previously invisible, to appear. The letter *Y* appeared if the correct answer was marked; the letter *N* appeared for incorrect answers.

Dependent Variables

Test Performance Measures

Three test performance measures were investigated. Ability level estimates were obtained using a Bayesian scoring procedure similar to the one developed by Owen (1975; see also McBride & Weiss, 1976, and Brown & Weiss, 1977, for applications of this ability estimation method). This scoring procedure provided a means of generating comparable scores for the conventional and adaptive tests. The posterior Bayesian variance, the second dependent variable used in this study, is the variance of the estimated ability score and can be considered an estimated standard error of estimate. The third dependent variable was the number of test questions omitted by each testee.

Psychological Reaction Scales

The psychological reactions to each condition were assessed by administering test reaction questions consisting of brief versions of four scales designed to assess reaction to knowledge of results, nervousness, motivation, and tendency to guess (see Betz & Weiss, 1976, for a description of the development of the scales from which these questions were selected). The test reaction questions are shown by scale in Appendix Table D along with the scaled scores used to obtain scores on the four scales. A student's score for each scale was the average of the scaled scores for the student's responses to the items in the scale.

The test reaction items were administered to each student twice, once after each test condition (computer-administered and paper-and-pencil). Students in the No-KR condition were given only the Nervousness, Motivation, and Guessing scales.

RESULTS

Test Characteristics

Test Items

To better interpret the meaning of any performance or motivational differences found between different testing conditions, it was important to examine the characteristics of the items administered under each testing condition. Because the computer-administered tests used a stradaptive strategy for item selection, it was not possible prior to administration to equate the item characteristics of the 20-item conventional paper-and-pencil tests to the 20-item computerized adaptive tests. As described earlier, items were divided between the paper-and-pencil and stradaptive item pools in order to equate, to the extent possible, item discriminations in the NBR condition and item bias in the BR condition.

Table 2 shows the mean, standard deviation, minimum and maximum values of item discrimination (*a*), difficulty (*b*), and bias parameters for the items in the conventional paper-and-pencil test and for the items actually administered in the computerized adaptive test under both BR and NBR conditions. For example, the average discrimination for items actually administered in the NBR adaptive test was 1.50, with discriminations of items administered ranging from about 1.00 to 2.27. These items also had a mean bias value of .72, indicating that the average item favored White students. For the conventional test in the NBR condition, the mean item discrimination was 1.57, with a range of 1.17 to 2.27.

Table 2
Item Discrimination (*a*), Difficulty (*b*), and Bias Values for the Conventional Paper-and-Pencil Tests and the Computerized Adaptive Tests in Bias-Reduced and Non-Bias-Reduced Conditions

Test and	Bias-Reduced (*N*=105)			Non-Bias-Reduced (*N*=106)		
Statistic	*a*	*b*	Bias	*a*	*b*	Bias
Conventional Test						
Mean	1.03	.02	-.05	1.57	.05	.83
S.D.	.47	.55	1.34	.28	.71	.36
Minimum	.09	-1.48	-5.46	1.17	-1.48	.22
Maximum	2.27	1.01	.74	2.27	1.46	1.71
Adaptive Test						
Mean	.84	-.10	-.20	1.50	-.41	.72
S.D.	.45	.91	1.22	.37	.59	.34
Minimum	.13	-1.61	-3.64	1.00	-1.51	.05
Maximum	1.96	2.04	1.29	2.27	.74	1.46

The data in Table 2 show that the strategy for item selection used in the BR condition did result in an adaptive test which was "bias-reduced," since the average bias value for items actually administered in the adaptive tests to students in the BR condition was -.20, which was lower than that for items administered in the NBR condition (mean = .72).

Not surprisingly, since NBR items were selected on the basis of their discrimination parameters, the average item administered in the adaptive test under the NBR condition was more discriminating (mean a=1.50) than the average item in the BR condition (mean a=.84). This was also reflected in the higher range of discrimination values in the NBR test.

In the conventional paper-and-pencil tests the item selection strategy resulted in the BR test having less "bias" against Black students (mean bias = -.05 compared to .83 in the NBR test), but it was also less discriminating (mean a=1.03 versus 1.57 for the NBR test). While the average item bias in the BR paper-and-pencil test favored Black students, examination of Appendix Table C indicates that this was attributable to a few items with large negative bias indices and that more of the items had small positive values of the bias index (i.e., favored White students). These items, however, had lower positive values of the bias index than most of the items in the NBR tests. Thus, while some of the items in the BR tests favored White students, the test items were, in general, more fair toward the Black students than the NBR tests.

Measurement Precision

Because increased item discrimination is related to increased item information, the NBR test might be expected to provide more precise ability estimates. In addition, previous research (Vale, 1975) has indicated that an adaptive test can yield more equiprecise measurement throughout the range of ability than a conventional peaked test. Using the Bayesian posterior variance as an estimate of the precision of measurement (Urry, 1977) at various levels of ability, Figures 2 and 3 provide support for both these expectations (numerical values for these figures are in Appendix Table E).

Figure 2 shows the mean Bayesian posterior variance for intervals of the Bayesian ability scores in the NBR condition; more precise measurement (lower posterior variance) was obtained with the adaptive test except for students whose ability level centered around the level of difficulty where the conventional test was peaked. In this range (θ=-.6 to .2) the conventional test had lower values of the Bayesian posterior variance.

Figure 3 shows the Bayesian posterior variance as a function of ability level for the BR condition for both adaptive and conventional tests. Under this test administration condition, items were selected by the adaptive test in order of their bias index, rather than by their discriminations. As Table 2 shows, the average discrimination of items administered in the adaptive test was lower than that in the conventional test. This is reflected in higher mean levels of the Bayesian posterior variance for the adaptive test for values of ability greater than θ=-1.00. In spite of this item selection procedure in the adaptive test, it still achieved lower average levels of the Bayesian posterior variance than did the conventional test for

Figure 2
Mean Bayesian Posterior Variance as a Function of Bayesian Ability Estimate for the Non-Bias-Reduced Adaptive and Conventional Tests

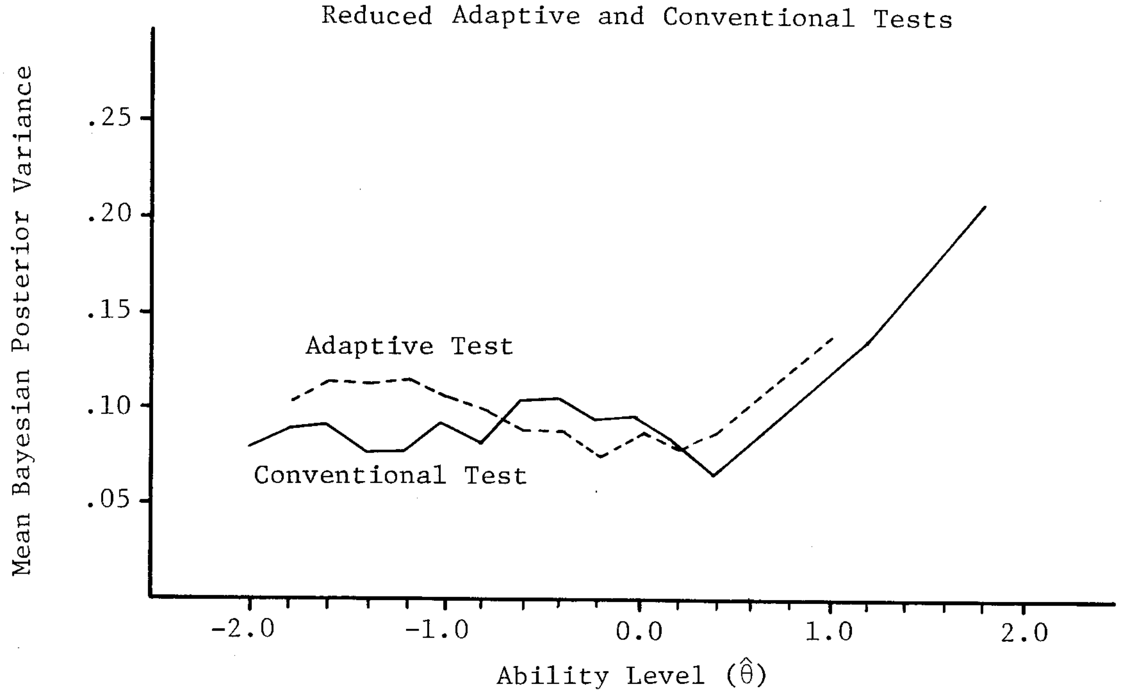
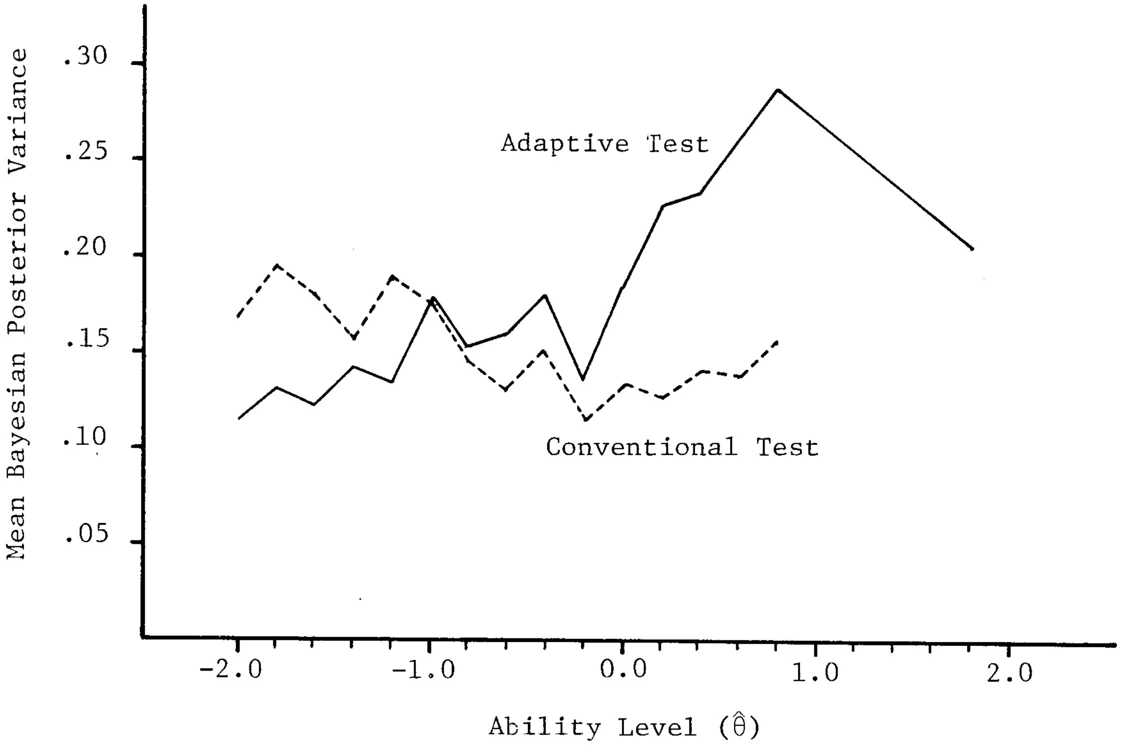


Figure 3
Mean Bayesian Posterior Variance as a Function of Bayesian Ability Estimate for the Bias-Reduced Conventional Test



ability levels less than $\theta=-1.00$. The adaptive test compared more favorably with the conventional test in the NBR condition (Figure 2), however, supporting earlier recommendations that items within strata should be selected by their discrimination values when using a stradaptive testing strategy (Weiss, 1974).

Dependent Variables

Test Performance Variables

Appendix Tables F, G, and H show the means and standard deviations of the Bayesian ability estimates, Bayesian posterior variances, and number of omitted responses, respectively, for all combinations of the independent variables. Appendix Table I contains the means and standard deviations of these three dependent variables for various combined groups.

Ability estimates. The results of the 2×2×2×2×2 repeated measures analysis of variance for the Bayesian ability estimates are shown in Table 3. As this table indicates, the only statistically significant ($p<.02$) main effect was for race, with White students scoring higher (means = -.61 and -.63 for the computerized adaptive and conventional paper-and-pencil tests, respectively; see Table I) than Black students (means = -.87 and -.85, respectively). The interpretation of this significant main effect must be qualified, however, by a marginally significant three-way interaction between Race, KR, and BR ($p=.07$) and a four-way interaction between Mode, Race, KR, and BR ($p<.06$).

Figure 4 shows the four-way interaction (since it subsumes the three-way interaction) graphically by separately plotting the three-way interactions for both the computerized and paper-and-pencil administration modes. From this figure it can be seen that Black students did best in both testing modes when the test was bias-reduced and no knowledge of results was provided (BR, No-KR). In both tests this condition eliminated the main effect for race which existed in the other conditions. Black students obtained lowest mean scores (-1.02) in the paper-and-pencil test (Figure 4a) when the test was bias-reduced and knowledge of results was provided (BR, KR). On the computer-administered test in this condition (Figure 4b), mean score for the Black students was also relatively low.

The four-way interaction appeared to result primarily from the differential effect of the administration conditions on mean scores of the White students. As Figure 4a shows, highest mean scores were obtained for the White students on the paper-and-pencil test under the NBR and No-KR conditions. On the adaptive test (Figure 4b), however, the White students obtained lowest mean scores under these conditions. Comparison of Figures 4a and 4b also shows a general tendency for the adaptive test to reduce mean differences due to the interaction of race and testing conditions, since for both racial groups there was less variability among mean ability level scores as a function of testing conditions for the adaptive test, despite higher score variability (see Appendix Tables E and I).

Consistency of ability estimates across modes. Of interest in comparing the computerized adaptive and conventional paper-and-pencil testing modes was

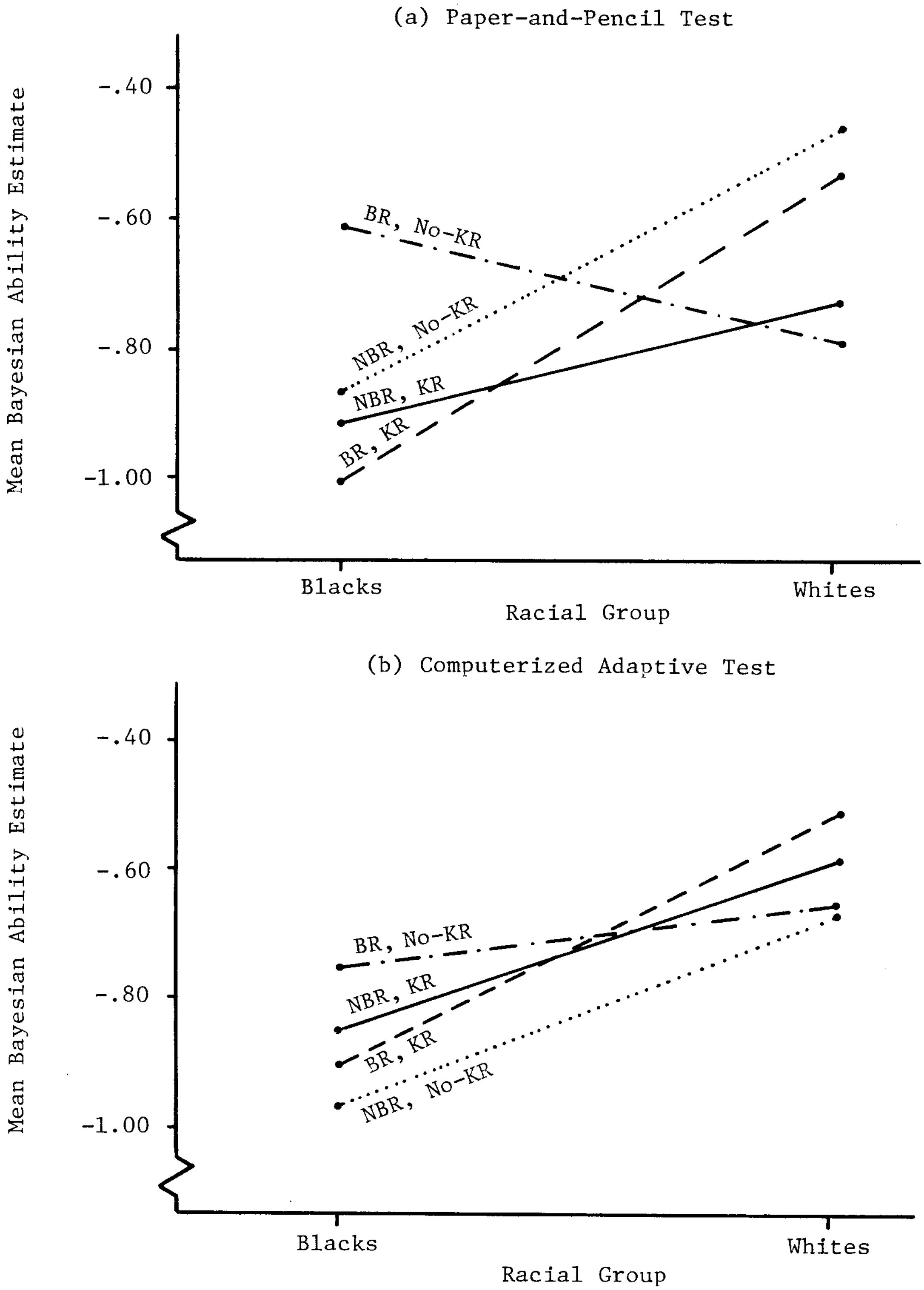
Table 3

Results of the Analysis of Variance for Bayesian Ability Estimates

Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	5.83	6.60	.001
Order	1	.95	1.08	.301
Knowledge of Results (KR)	1	.17	.19	.663
Bias Reduction (BR)	1	.09	.10	.754
Two-Way Interactions				
Race × Order	1	1.04	1.17	.280
Race × KR	1	.66	.75	.388
Order × KR	1	.00	.00	.982
Race × BR	1	.36	.41	.521
Order × BR	1	.19	.22	.640
KR × BR	1	.01	.02	.897
Three-Way Interactions				
Race × Order × KR	1	.01	.01	.910
Race × Order × BR	1	1.23	1.39	.240
Race × KR × BR	1	2.93	3.31	.070
Order × KR × BR	1	.49	.56	.456
Four-Way Interaction				
Race × Order × KR × BR	1	.51	.57	.450
Error	199	.88		
Within Subjects				
Main Effect				
Mode	1	.00	.02	.876
Two-Way Interactions				
Mode × Race	1	.14	.92	.338
Mode × Order	1	.10	.68	.409
Mode × KR	1	.43	2.83	.094
Mode × BR	1	.02	.17	.682
Three-Way Interactions				
Mode × Race × Order	1	.01	.08	.774
Mode × Race × KR	1	.21	1.40	.238
Mode × Order × KR	1	.01	.07	.799
Mode × Race × BR	1	.02	.15	.698
Mode × Order × BR	1	.43	2.84	.093
Mode × KR × BR	1	.14	.94	.333
Four-Way Interactions				
Mode × Race × Order × KR	1	.05	.30	.583
Mode × Race × Order × BR	1	.06	.39	.532
Mode × Race × KR × BR	1	.56	3.65	.057
Mode × Order × KR × BR	1	.44	2.86	.092
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.20	1.30	.255
Error	199	.15		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 4
Four-way Interaction of Mode of Administration, Race, Knowledge of Results (KR), and Bias Reduction (BR) for Bayesian Ability Estimates



(a) Paper-and-Pencil Test

(b) Computerized Adaptive Test

the equivalence of the ability estimates obtained from the computerized and paper-and-pencil administrations. While the analyses of variance examined group level effects of test mode, it is also relevant to examine the similarity of rank orderings of individual student ability estimates across the two modes of test administration.

Pearson product-moment correlations between the ability estimates from the computer-administered adaptive test and the conventional paper-and-pencil test indicated substantial, but far from perfect, agreement between the two estimates for the sample as a whole (r=.73), for Black students (r=.70), for White students (r=.74), for students taking the BR tests (r=.72), and for students taking the NBR tests (r=.73). These correlations were all significantly different from zero (p<.01), but did not differ significantly from each other.

One probable reason for the moderate level of similarity of the ability estimates in the two modes of administration relates to the adaptive nature of the computer-administered tests. The distribution of students falling into various ability level intervals (see Appendix Table E), and the larger standard deviation of ability estimates in the adaptive test (*S.D.* = .80) as compared to the paper-and-pencil test (*S.D.* = .63), indicate that the adaptive test spread students out more on the ability continuum than did the conventional test. While ICC theory suggests that using Bayesian scoring ability estimates should not be dependent on the difficulty level of the items given, it appears that the peaked paper-and-pencil test was not able to locate people as well on the ability continuum if their ability levels were not near the point at which the test was peaked.

Bayesian posterior variance. Table 4 shows the results of the five-way repeated measures analysis of variance for the Bayesian posterior variance scores. A highly significant (p<.01) main effect for the bias-reduction factor was found, indicating that errors of measurement were larger in the BR tests (see Table I). This is consistent with the greater average discrimination of items in the NBR tests. The data in Table I also show that for the NBR tests, in which the adaptive test selected available items which were most discriminating, the adaptive test provided more precise ability estimates than the paper-and-pencil tests. For the BR tests, there was no advantage of the adaptive test over the paper-and-pencil tests in terms of accuracy of ability estimates. The bias-reduction factor was also involved, however, in the significant Race×Order×BR, Mode×BR, and Mode×Order×BR interactions. In addition, a significant Mode×Order effect was found.

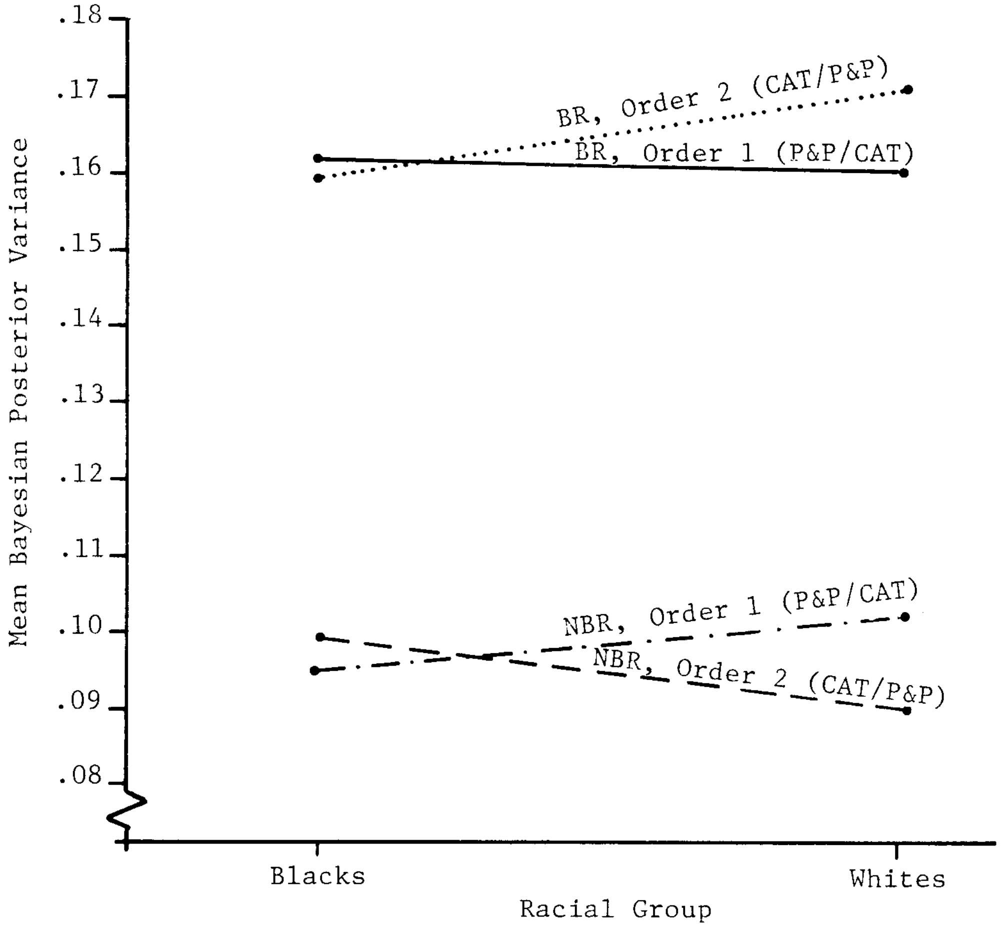
Figure 5 shows the Race×Order×BR three-way interaction. The increased precision obtained in the NBR conditions is clear in this figure, since lower values of the Bayesian posterior variance were obtained with the more highly discriminating items. The figure also shows that for the White group, posterior variances in the BR tests were smaller when the paper-and-pencil test was administered first (BR, P&P/CAT), while posterior variances were smaller in the NBR tests when the adaptive test was administered first (NBR, CAT/P&P). This pattern was reversed for Black students. In addition, the testing conditions had a greater effect on the Bayesian posterior variances for the White students.

Table 4
Results of the Analysis of Variance for Bayesian Posterior Variance Scores

Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	.00	.35	.554
Order	1	.00	.71	.401
Knowledge of Results (KR)	1	.00	3.06	.082
Bias Reduction (BR)	1	.45	582.28	.001
Two-Way Interactions				
Race × Order	1	.00	.066	.798
Race × KR	1	.00	1.58	.210
Order × KR	1	.00	.02	.893
Race × BR	1	.00	.67	.414
Order × BR	1	.00	2.78	.097
KR × BR	1	.00	2.30	.131
Three-Way Interactions				
Race × Order × KR	1	.00	3.47	.064
Race × Order × BR	1	.00	4.69	.031
Race × KR × BR	1	.00	.33	.564
Order × KR × BR	1	.00	.03	.870
Four-Way Interaction				
Race × Order × KR × BR	1	.00	.15	.697
Error	199	.00		
Within Subjects				
Main Effect				
Mode	1	.00	.50	.478
Two-Way Interactions				
Mode × Race	1	.00	2.16	.143
Mode × Order	1	.02	13.17	.001
Mode × KR	1	.00	.03	.872
Mode × BR	1	.01	6.01	.015
Three-Way Interactions				
Mode × Race × Order	1	.00	1.50	.222
Mode × Race × KR	1	.00	.16	.691
Mode × Order × KR	1	.00	.62	.430
Mode × Race × BR	1	.00	.67	.413
Mode × Order × BR	1	.01	9.32	.003
Mode × KR × BR	1	.00	1.74	.188
Four-Way Interactions				
Mode × Race × Order × KR	1	.00	.54	.463
Mode × Race × Order × BR	1	.00	2.12	.147
Mode × Race × KR × BR	1	.00	.00	.959
Mode × Order × KR × BR	1	.00	.12	.725
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.00	.04	.849
Error	199	.00		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 5
Three-Way Interaction of Race, Order of Administration, and Bias-Reduction (BR) for Bayesian Posterior Variance

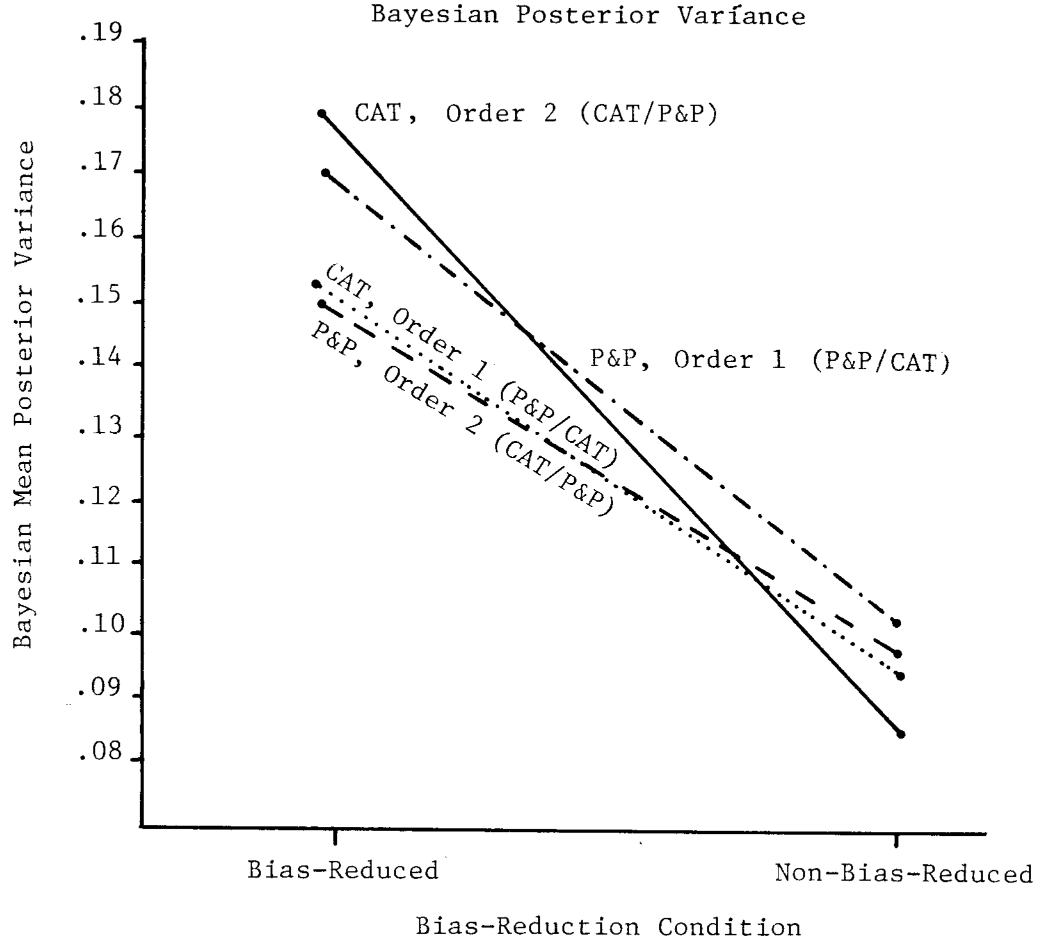


All the other significant interactions for the Bayesian posterior variance measure were subsumed in the Mode×Order×BR three-way interaction shown in Figure 6.

These data show that the combination of bias-reduced administration and order of administration affected Bayesian posterior variances on the adaptive test. Specifically, when the adaptive test was administered first (Order 2), it had the highest average level of the posterior variance among all test administration conditions in the BR condition and the lowest level in the NBR condition.

Number of omitted responses. Table 5 shows the results of the analysis of variance for the number of omitted responses. These data indicate a statistically significant (p<.02) main effect for KR,with students omitting more responses when KR was not given (see Table I). Examination of the statistically significant (p<.05) two-way interaction of the KR variable with the race factor (see Figure 7), however, indicates that this effect of KR on the number of omitted responses was largely due to its effect on the Black students.

Figure 6
Three-Way Interaction of Mode of Administration,
Order of Administration, and Bias-Reduction (BR) for
Bayesian Posterior Variance



As Figure 7 indicates, KR had a differential effect on the Black students, but no effect on the White students. When KR was administered, Black students omitted fewer items (mean = 1.88) than when KR was not given (mean = 3.89). In comparison, White students omitted an average of 2.75 and 2.68 items under KR and No-KR conditions, respectively.

The only other statistically significant interaction for omitted responses was the three-way interaction of Mode×Race×Order ($p<.05$). This interaction, pictured in Figure 8, shows that Black and White students differed in the relative number of responses they omitted on the paper-and-pencil test depending on whether that test was taken first or second. For the Black students, the highest mean number of omitted responses as a group occurred when the paper-and-pencil test was taken second (Order 2); and the fewest, when this test was taken first (Order 1). For the White students, the mean number of omitted responses on the paper-and-pencil test was highest when this test was given first (Order 1) and fewest when this test was given second (Order 2). In addition, the test administration variables resulted in slightly greater mean differences for the White students than for the Black students.

Table 5

Results of the Analysis of Variance for Number of Omitted Responses

	Degrees of	Mean		
Source of Variation	Freedom	Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	.25	.01	.914
Order	1	14.08	.65	.419
Knowledge of Results (KR)	1	123.64	5.76	.017
Bias Reduction (BR)	1	7.20	.33	.563
Two-Way Interactions				
Race × Order	1	14.93	.69	.405
Race × KR	1	99.18	4.62	.033
Order × KR	1	69.35	3.23	.074
Race × BR	1	.00	.00	.993
Order × BR	1	8.00	.37	.542
KR × BR	1	39.93	1.86	.174
Three-Way Interactions				
Race × Order × KR	1	.75	.03	.852
Race × Order × BR	1	3.90	.18	.671
Order × KR × BR	1	37.89	1.76	.186
Four-Way Interaction				
Race × Order × KR × BR	1	1.72	.08	.777
Error	206	21.48		
Within Subjects				
Main Effect				
Mode	1	.48	.06	.811
Two-Way Interactions				
Mode × Race	1	.47	.06	.813
Mode × Order	1	.02	.00	.956
Mode × KR	1	10.33	1.24	.267
Mode × BR	1	.26	.03	.859
Three-Way Interactions				
Mode × Race × Order	1	38.53	4.63	.033
Mode × Race × KR	1	23.71	2.85	.093
Mode × Order × KR	1	5.71	.68	.409
Mode × Race × BR	1	5.34	.64	.424
Mode × Order × BR	1	18.04	2.17	.143
Mode × KR × BR	1	2.34	.28	.596
Four-Way Interactions				
Mode × Race × Order × KR	1	.45	.05	.816
Mode × Race × Order × BR	1	8.70	1.04	.308
Mode × Race × KR × BR	1	22.37	2.69	.103
Mode × Order × KR × BR	1	3.68	.44	.507
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.38	.05	.830
Error	206	8.32		

*Estimated probability of error in rejection of the null hypothesis of no mean difference.

Figure 7
Two-Way Interaction of Race and Knowledge of Results (KR) for Number of Omitted Responses

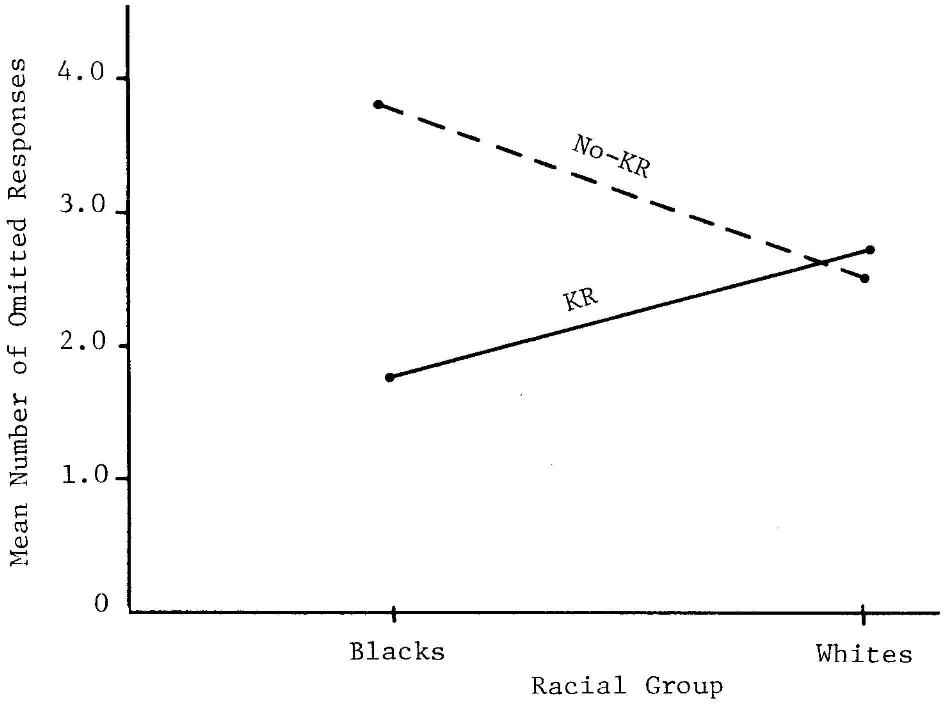
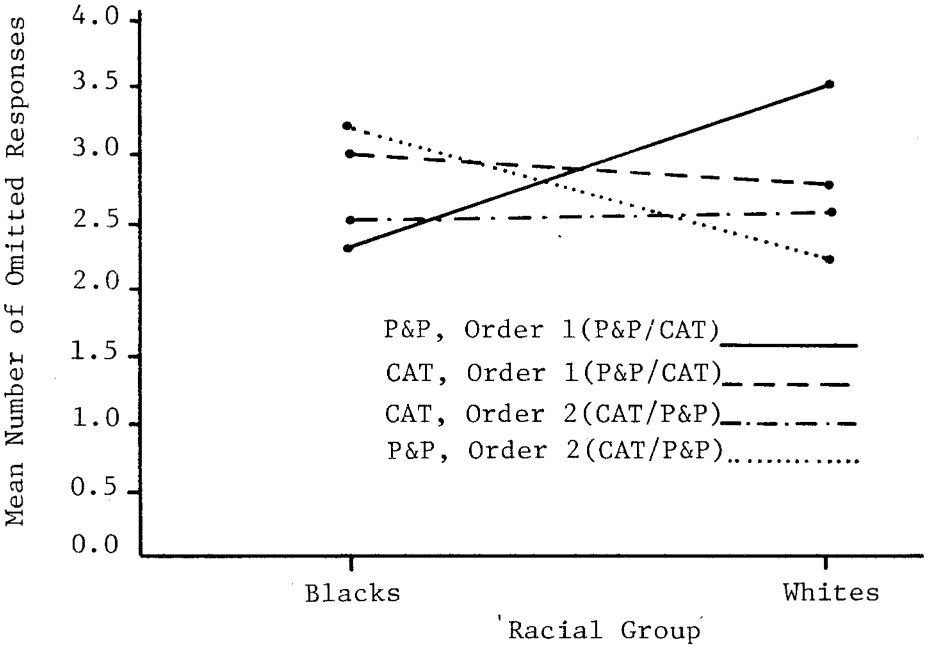


Figure 8
Three-Way Interaction of
Mode of Administration, Race, and Order of Administration
for Number of Omitted Responses



Psychological Reaction Variables

Means and standard deviations of the psychological reactions scales for all experimental conditions are in Appendix Tables J, K, L, and M, respectively, for the Knowledge of Results, Nervousness, Motivation, and Guessing scales. The means and standard deviations of the four psychological test reactions scales for the combined Racial, Bias-Reduction, Knowledge of Results, Order of Administration, and Mode of Administration groups are given in Appendix Table N.

Knowledge of Results. Table 6 gives the results of the analysis of variance of the scores on the reaction to Knowledge of Results scale. There was a statistically significant (*p*=.001) effect for race in the ANOVA of the reaction to Knowledge of Results scores, with Black students scoring higher on this scale than White students. This indicated a more negative attitude toward receiving KR after each item on the part of the Black students, i.e., they were more inclined to report that receiving KR made them nervous and interfered with their concentration.

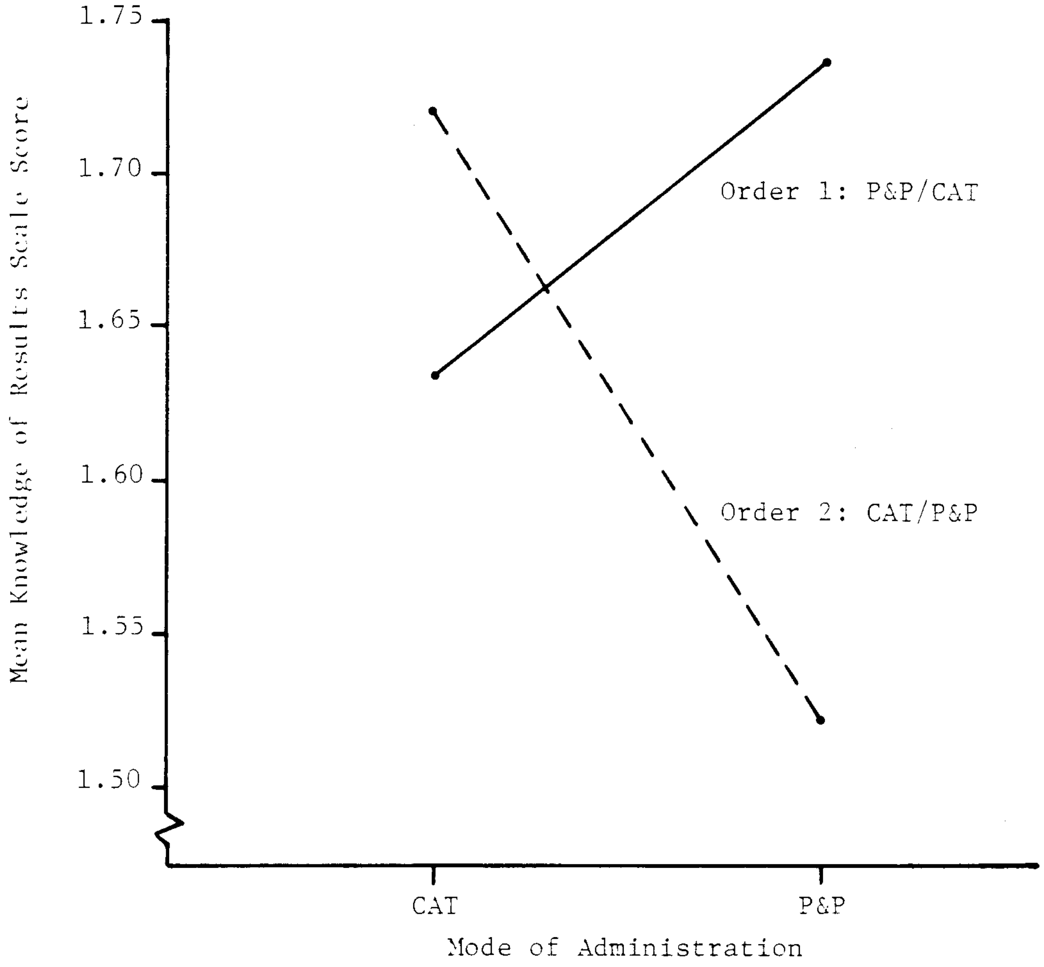
Table 6
Results of the Analysis of Variance of the Knowledge of Results Scale Scores

Source of Variance	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	11.25	12.59	.001
Order	1	.39	.44	.509
Bias Reduction (BR)	1	.06	.06	.804
Two-Way Interactions				
Race × Order	1	.96	1.08	.302
Race × BR	1	.10	.11	.742
Order × BR	1	1.24	1.39	.242
Three-Way Interaction				
Race × Order × BR	1	.92	1.03	.313
Error	88	.89		
Within Subjects				
Main Effect				
Mode	1	.29	1.42	.236
Two-Way Interactions				
Mode × Race	1	.00	.02	.899
Mode × Order	1	.94	4.63	.034
Mode × BR	1	.07	.35	.558
Three-Way Interactions				
Mode × Race × Order	1	.27	1.31	.256
Mode × Race × BR	1	.48	2.34	.129
Mode × Order × BR	1	.23	1.13	.290
Four-Way Interaction				
Mode × Race × Order × BR	1	.19	.91	.342
Error	88	.20		

*Estimated probability of error in rejecting the null hypothesis of no difference in group means.

The Mode×Order interaction was also statistically significant ($p<.05$) and is illustrated in Figure 9. Students reported a more favorable attitude toward KR (i.e., lower mean scale scores) during the second test than during the first. This was particularly true when the paper-and-pencil test was administered second, which was the condition resulting in the most favorable reactions to KR. The data in Figure 9 also show that students' reactions to computer-administered KR were less affected by the order of its administration than was paper-and-pencil-administered KR.

Figure 9
Two-Way Interaction of Mode of Administration and Order of Administration for the Knowledge of Results Scale Scores



Nervousness. The means and standard deviations of responses on the Nervousness scale are reported in Appendix Tables K and N; Table 7 gives the results of the analysis of variance for this scale.

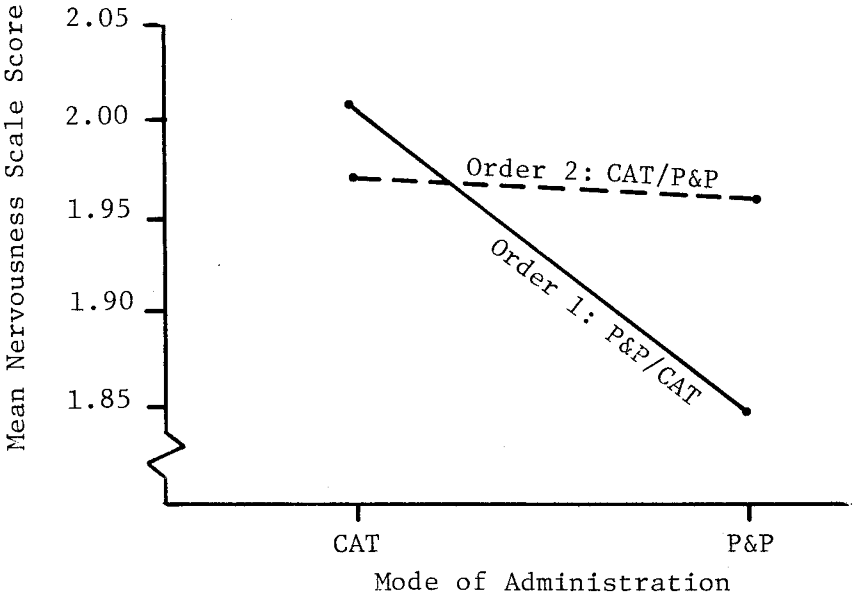
The only main effect that emerged as statistically significant ($p<.05$) was that of mode of administration, in which students reported that they were more nervous while taking the computer-administered test (mean = 2.02; see Table N) than they were while taking the paper-and-pencil test (mean = 1.91). The Mode×Order interaction was marginally significant ($p=.076$) and is shown in Figure 10. This figure shows that students reported lowest levels of nervousness

Table 7
Results of the Analysis of Variance of Nervousness Scale Scores

Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	.65	.90	.344
Order	1	.01	.01	.909
Knowledge of Results (KR)	1	1.58	2.19	.140
Bias Reduction (BR)	1	.60	.83	.364
Two-Way Interactions				
Race × Order	1	1.34	1.86	.174
Race × KR	1	.18	.25	.619
Race × BR	1	1.44	2.00	.159
Order × KR	1	.14	.19	.663
Order × BR	1	5.37	7.45	.007
KR × BR	1	.57	.79	.376
Three-Way Interactions				
Race × Order × KR	1	1.01	1.40	.238
Race × Order × BR	1	.38	.53	.468
Race × KR × BR	1	.27	.38	.540
Order × KR × BR	1	.09	.13	.717
Four-Way Interaction				
Race × Order × KR × BR	1	3.17	4.41	.037
Error	185	.72		
Within Subjects				
Main Effect				
Mode	1	1.22	5.01	.026
Two-Way Interactions				
Mode × Race	1	.12	.50	.480
Mode × Order	1	.78	3.19	.076
Mode × KR	1	.07	.30	.584
Mode × BR	1	.87	3.57	.060
Three-Way Interactions				
Mode × Race × Order	1	.01	.03	.868
Mode × Race × KR	1	.01	.05	.825
Mode × Race × BR	1	.16	.66	.416
Mode × Order × KR	1	.14	.56	.455
Mode × Order × BR	1	.16	.64	.423
Mode × KR × BR	1	.01	.02	.825
Four-Way Interactions				
Mode × Race × Order × KR	1	.00	.00	.985
Mode × Race × Order × BR	1	.15	.61	.435
Mode × Race × KR × BR	1	.02	.09	.760
Mode × Order × KR × BR	1	.04	.16	.689
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.30	1.25	.265
Error	185	.24		

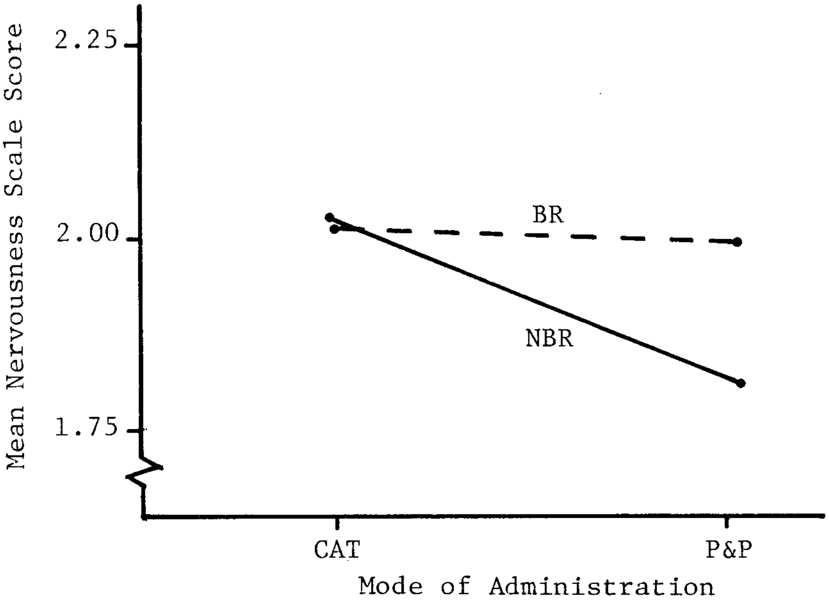
*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Figure 10
Two-Way Interaction of Mode of Administration and
Order of Administration for Nervousness Scale Scores



when the paper-and-pencil test (Order 1) was administered first in a pair of tests (mean = 1.85) and highest levels when they were subsequently transferred to the computerized adaptive test (mean = 2.06). However, when students were first administered the computerized test (Order 2), their reported levels of nervousness remained about the same across both tests.

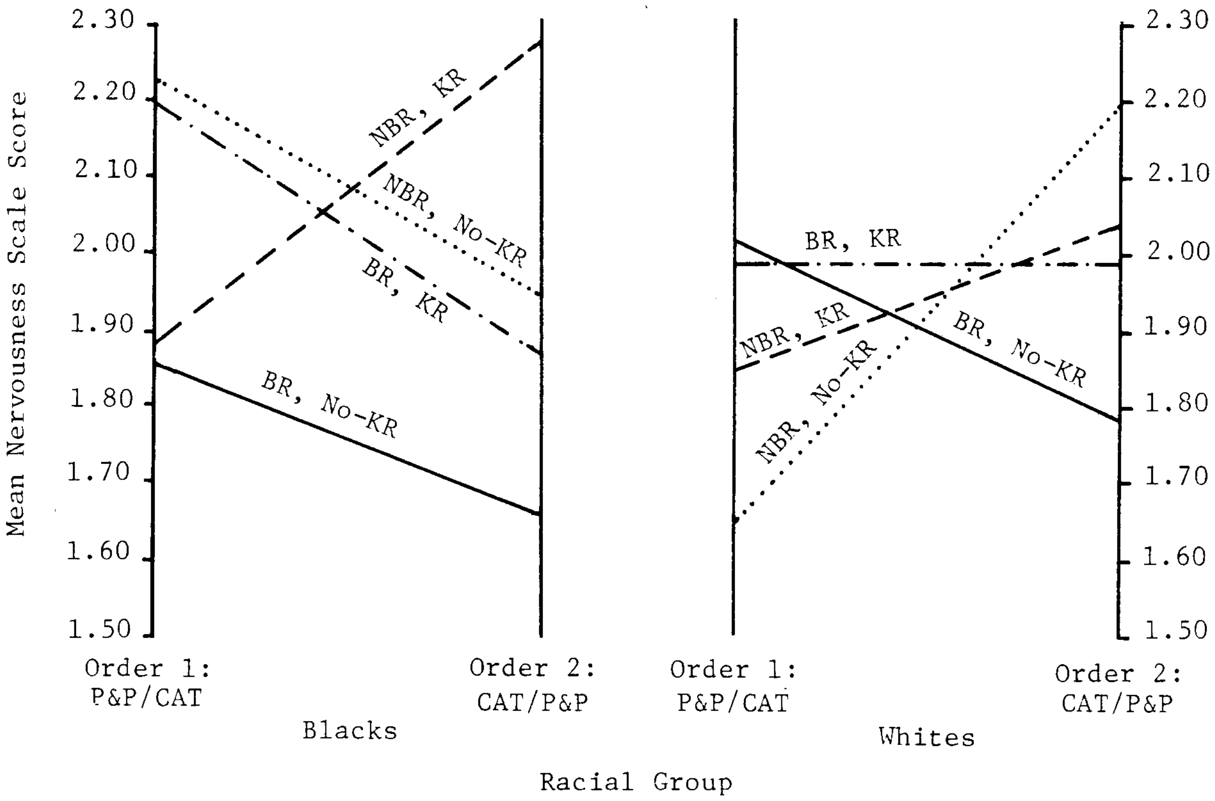
Figure 11
Two-Way Interaction of Mode of Administration
and Bias-Reduction (BR) for Nervousness Scale Scores



The Mode×BR interaction was also marginally significant (p=.06). Inspection of the graph of this interaction (Figure 11) indicates that the students reported equal levels of nervousness in both BR and NBR tests when they were administered adaptively by computer. When tests were administered by paper-and-pencil, however, lower levels of nervousness were observed in the BR condition.

There was also a statistically significant (p=.007) Order×BR interaction. Interpretation of this interaction is complicated by the presence of a four-way Race×Order×BR×KR interaction (p=.037), which is shown in Figure 12. As Figure 12 shows, reported nervousness of Black and White students was differentially affected by the Order, KR, and BR test administration conditions. Black students reported lower levels of nervousness when the computerized adaptive test was administered first if the tests were administered in the BR mode (with or without KR) and when the NBR test was administered without KR; they reported highest levels of nervousness when the NBR adaptive test was administered first with KR. For the Black students, lowest levels of nervousness were reported in the BR, No-KR condition, regardless of test order. For the White students, order of administration did not affect their reported nervousness in the BR, KR condition; the NBR, No-KR condition resulted in lowest levels of reported nervousness when the paper-and-pencil test was

Figure 12
Four-Way Interaction of Race, Order of Administration, Knowledge of Results (KR), and Bias-Reduction (BR) for Nervousness Scale Scores



administered first and highest levels of nervousness when it was administered second. Order of administration also affected the White students in opposite ways under the other two test administration condition combinations.

Motivation. The means and standard deviations of responses on the Motivation scale are given in Appendix Tables L and N; results of the analysis of variance for this scale are given in Table 8. Again, there was a statistically significant (p<.01) main effect for mode of administration, with students reporting that they were more motivated to perform well when they were taking the computer-administered test (mean = 2.99; see Table N) than when they took the paper-and-pencil test (mean = 2.86).

The Mode×Order interaction was marginally significant (p=.071) for this scale, but it was subsumed in the significant (p=.022) four-way Mode×Order×Race×BR interaction. The two-way Mode×BR and Race×Order interactions were also statistically significant (p=.005 and .021, respectively); these were also subsumed in the significant Race×Order×Mode×BR interaction, which is shown in Figure 13.

Figure 13
Four-Way Interaction of Mode of Administration, Race, Order of Administration, and Bias-Reduction (BR) for Motivation Scale Scores

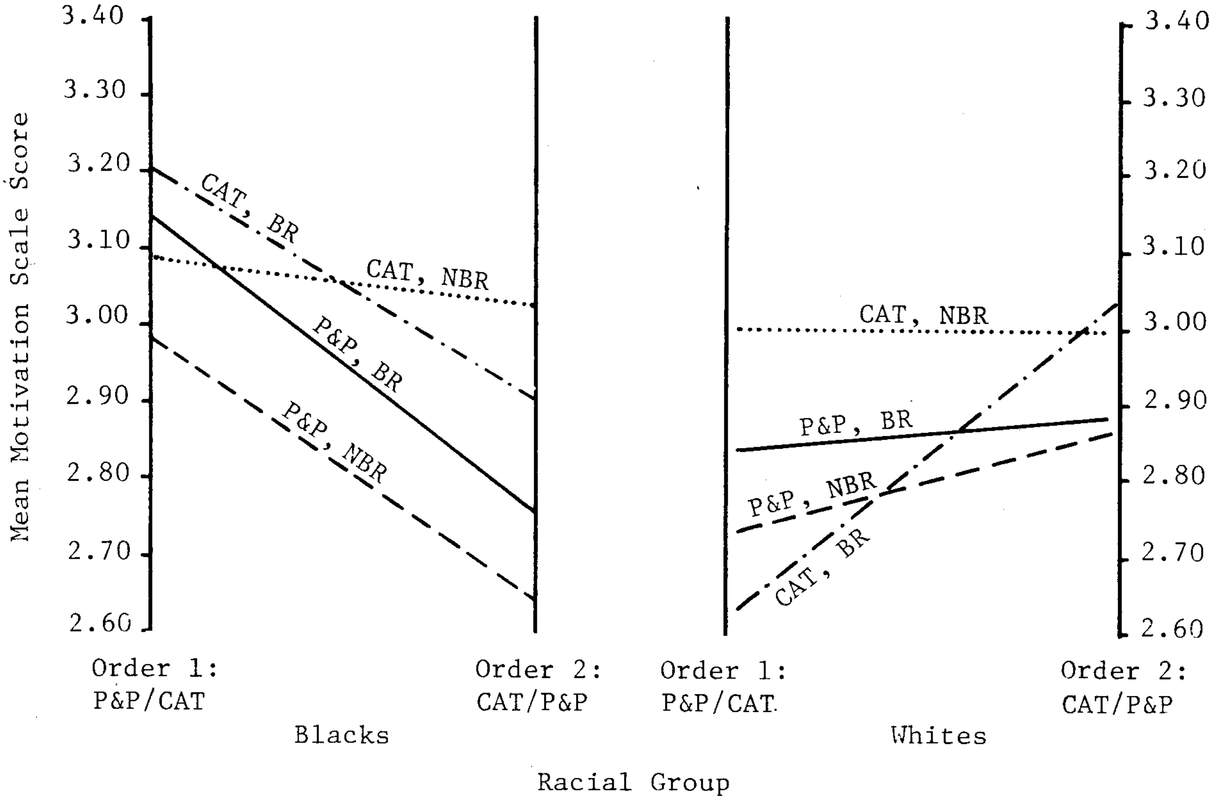


Figure 13 shows that reported motivation was uniformly lower for Black students in Order 2 (CAT/P&P) than in Order 1 (P&P/CAT). However, Order 2 had a greater effect on motivation reported after the paper-and-pencil test administration than after administration of the adaptive test. For the Black

Table 8

Results of the Analysis of Variance for Motivation Scale Scores

Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	1.00	1.16	.283
Order	1	.22	.25	.616
Knowledge of Results (KR)	1	.20	.23	.628
Bias Reduction (BR)	1	.00	.00	.997
Two-Way Interactions				
Race × Order	1	4.68	5.41	.021
Race × KR	1	.21	.24	.624
Race × BR	1	.29	.34	.562
Order × KR	1	.79	.91	.340
Order × BR	1	.02	.03	.867
KR × BR	1	.16	.19	.665
Three-Way Interactions				
Race × Order × KR	1	1.69	1.96	.164
Race × Order × BR	1	.39	.44	.506
Race × KR × BR	1	.92	1.07	.303
Order × KR × BR	1	2.34	2.70	.102
Four-Way Interaction				
Race × Order × KR × BR	1	5.10	5.89	.016
Error	185	.87		
Within Subjects				
Main Effect				
Mode	1	2.17	14.04	.000
Two-Way Interactions				
Mode × Race	1	.09	.59	.445
Mode × Order	1	.51	3.31	.071
Mode × KR	1	.25	1.64	.202
Mode × BR	1	1.25	8.09	.005
Three-Way Interactions				
Mode × Race × Order	1	.02	.16	.689
Mode × Race × KR	1	.31	2.01	.158
Mode × Race × BR	1	.21	1.33	.251
Mode × Order × KR	1	.25	1.60	.208
Mode × Order × BR	1	.13	.84	.360
Mode × KR × BR	1	.01	.08	.783
Four-Way Interactions				
Mode × Race × Order × KR	1	.17	1.10	.297
Mode × Race × Order × BR	1	.83	5.35	.022
Mode × Race × KR × BR	1	.01	.04	.833
Mode × Order × KR × BR	1	.38	2.47	.118
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.03	.20	.654
Error	185	.15		

*Estimated probability of error in rejecting the null hypothesis of no mean differences.

students, lowest levels of motivation in both orders of administration were reported for the NBR paper-and-pencil test; highest levels of reported motivation were reported in Order 1 on the BR adaptive test. In general, order of administration had an opposite effect on White students; reported levels of motivation were higher for Order 2 than for Order 1. For Whites, the BR adaptive test resulted in lowest levels of reported motivation when it was administered second and highest levels when it was administered first. For both the Black and White groups, the NBR adaptive test was the only testing condition for which order of administration did not affect reported motivation.

Guessing. The means and standard deviations of responses on the Guessing scale are reported in Appendix Tables M and N, and the results of the analysis of variance for that scale are given in Table 9.

Figure 14
Three-Way Interaction of Mode of Administration, Race, and Order of Administration for Guessing Scale Scores

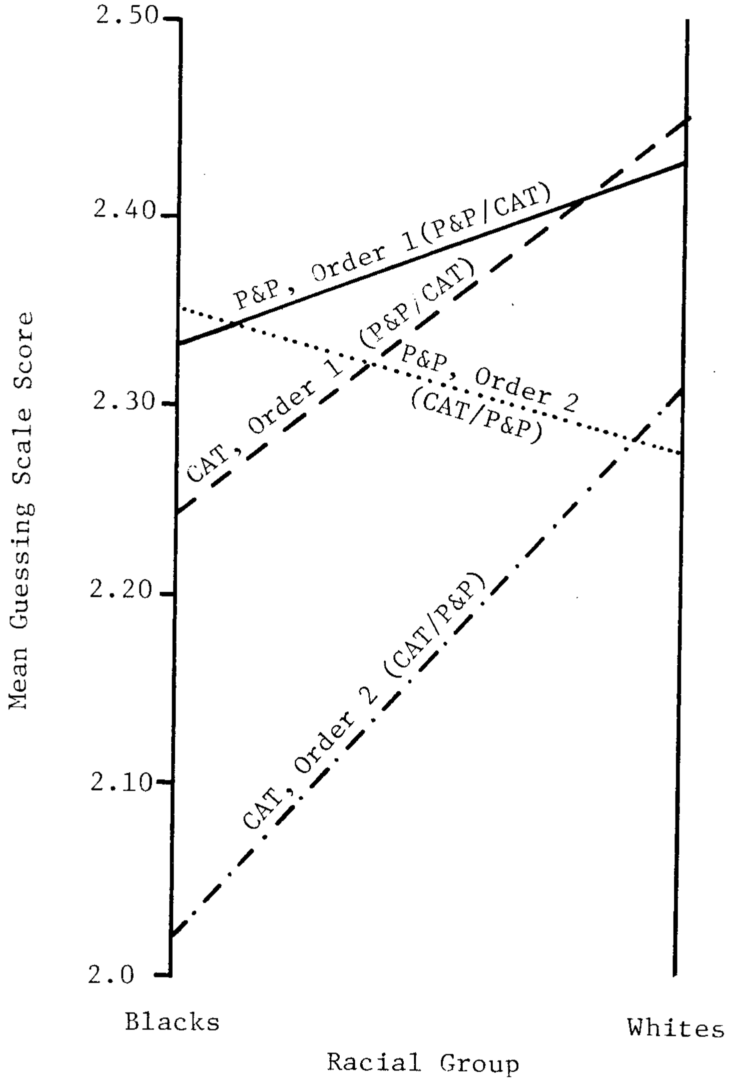


Table 9

Results of the Analysis of Variance for Guessing Scale Scores

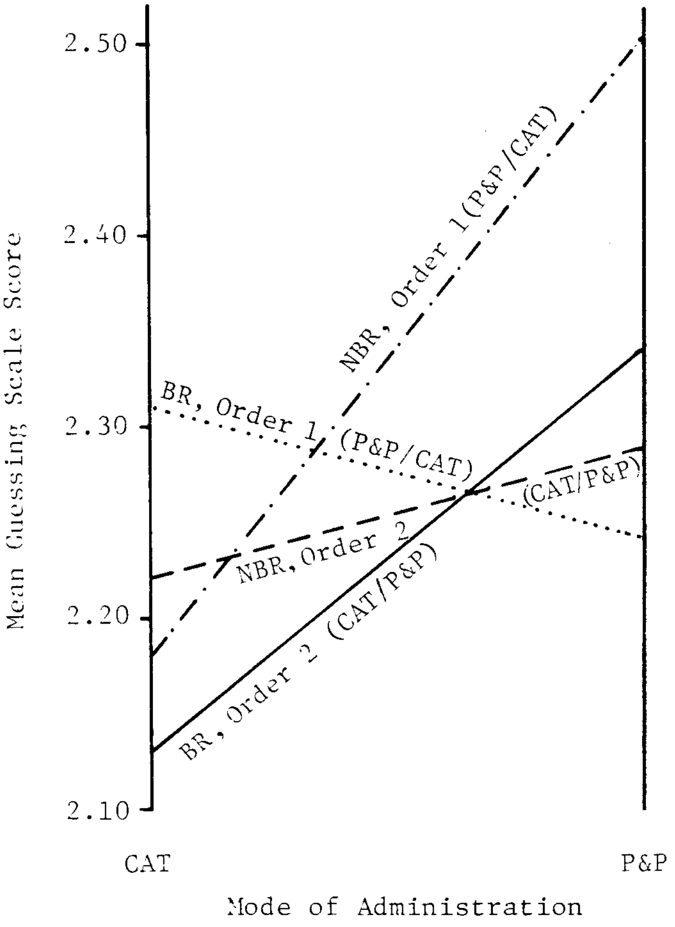
Source of Variation	Degrees of Freedom	Mean Square	*F*	*p**
Between Subjects				
Main Effects				
Race	1	.47	.63	.429
Order	1	.51	.68	.412
Knowledge of Results (KR)	1	.10	.14	.710
Bias Reduction (BR)	1	.33	.44	.507
Two-Way Interactions				
Race × Order	1	.00	.00	.953
Race × KR	1	1.73	2.31	.130
Race × BR	1	.05	.06	.800
Order × KR	1	.21	.28	.596
Order × BR	1	.18	.24	.627
KR × BR	1	.25	.34	.562
Three-Way Interactions				
Race × Order × KR	1	1.34	1.79	.183
Race × Order × BR	1	.22	.29	.591
Race × KR × BR	1	.17	.22	.636
Order × KR × BR	1	2.74	3.67	.057
Four-Way Interaction				
Race × Order × KR × BR	1	.47	.63	.427
Error	185	.75		
Within Subjects				
Main Effect				
Mode	1	2.06	6.06	.015
Two-Way Interactions				
Mode × Race	1	.34	1.01	.316
Mode × Order	1	.00	.01	.936
Mode × KR	1	.04	.11	.739
Mode × BR	1	.85	2.52	.114
Three-Way Interactions				
Mode × Race × Order	1	1.26	3.72	.055
Mode × Race × KR	1	.17	.50	.480
Mode × Race × BR	1	.03	.08	.781
Mode × Order × KR	1	.27	.79	.375
Mode × Order × BR	1	1.53	4.51	.035
Mode × KR × BR	1	.01	.04	.838
Four-Way Interactions				
Mode × Race × Order × KR	1	.06	.19	.664
Mode × Race × Order × BR	1	.00	.00	.963
Mode × Race × KR × BR	1	.01	.04	.850
Mode × Order × KR × BR	1	.22	.65	.421
Five-Way Interaction				
Mode × Race × Order × KR × BR	1	.38	1.13	.290
Error	185	.34		

*Estimated probability of error in rejection of the null hypothesis of no mean differences.

Again, there was a statistically significant (p<.02) main effect for mode of administration, with all students reporting that they guessed more often on the conventional paper-and-pencil tests (mean = 2.34) than on the computer-administered adaptive tests (mean = 2.21; see Table N). The interpretation of this difference was complicated by the significant Mode×Race×Order interaction (p=.055), shown in Figure 14. In three of the four Mode×Order conditions, the Black students reported that they guessed less than did the White students. Lowest levels of guessing were reported by Black students on the computer-administered adaptive tests, particularly when the computerized test was administered first (Order 2). White students reported highest levels of guessing, on both the adaptive and paper-and-pencil tests, when the paper-and-pencil test was administered first (Order 1); they reported lowest levels of guessing on both tests when the adaptive test was administered first (Order 2).

The three-way Mode×Order×BR interaction was also statistically significant (p=.035) and is shown in Figure 15. The highest level of guessing was reported on the NBR paper-and-pencil test when it was administered first (Order 1). Lowest levels of guessing were reported on the BR adaptive test

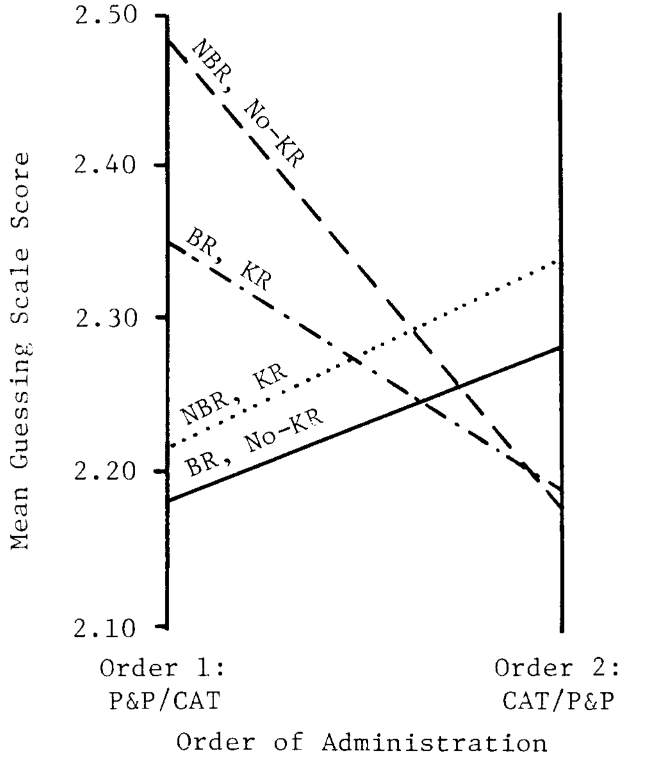
Figure 15
Three-Way Interaction of Mode of Administration,
Order of Administration, and Bias-Reduction (BR)
for Guessing Scale Scores



when it was administered first (Order 2). For three of the four comparisons between the adaptive and conventional tests, lower levels of guessing were reported on the adaptive tests; the exception was the BR adaptive test when it was administered first in the pair (Order 1).

The three-way Order×KR×BR interaction was also marginally significant (p=.057); Figure 16 shows the mean guessing scores for these test administration conditions. Highest levels of guessing were reported under Order 1 (P&P/CAT) when the NBR test was administered without KR; when the same test was administered under the reverse order, lowest levels of guessing were reported.

Figure 16
Three-Way Interaction of Order of Administration, Knowledge of Results (KR), and Bias-Reduction (BR) for Guessing Scale Scores



Relationship Between Ability Estimates and Psychological Reactions

Table 10 shows the Pearson product-moment correlations between the Bayesian ability estimates for the conventional paper-and-pencil and computerized adaptive tests and corresponding scores on the psychological reaction scales for each test. These data show that the only psychological variable which was not related to ability scores was reported motivation. There was a small to moderate tendency for students who performed better on the tests to be less nervous (r=-.25 for the paper-and-pencil test; r=-.16 for the adaptive test) and to report less tendency to guess (r=-.30) for the paper-and-pencil test. The strongest relationship was between ability scores and students' reactions to knowledge of results. Higher ability students felt better about receiving KR (*r*=-.44 for the paper-and-pencil test; *r*=-.38 for the adaptive test) than lower ability students. This is not surprising in the paper-and-pencil test, where lower ability students would receive more negative feedback on their performance; but the effect also held for the adaptive test, which should have provided comparable amounts of positive and negative feedback for high- and low-ability students. In all cases where the psychological variables related to the ability scores (i.e., nervousness, reaction to knowledge of results, and guessing), the relationship between these variables was stronger in the conventional paper-and-pencil test than in the computerized adaptive test. This may indicate a "homogenizing" effect on students' reactions to testing when tests are administered adaptively by computer.

Table 10
Correlations of Bayesian Ability Estimates on the Conventional Paper-and-Pencil (P&P) and Computerized Adaptive Tests (CAT) with Psychological Reactions Scale Scores

Test	Nervousness	Knowledge of Results (KR)	Motivation	Guessing
P&P	-.25**	-.44**	.03	-.30**
CAT	-.16*	-.38**	-.04	-.05

**p*<.05; ***p*<.01

DISCUSSION

The results indicate that the bias-reduced strategy of test construction used in this study to reduce racial performance differences was partially successful. Although the BR tests contained some items which clearly favored Black students, the majority of the items represented only a reduction in the degree to which the items favored White students over the NBR tests. In general, the White students obtained higher ability estimates than the Black students. However, mean ability estimates for the Black students were comparable to those of the White students on both the conventional paper-and-pencil and computerized adaptive tests when the bias-reduced tests were given without the provision of KR. When KR was provided on the BR tests, Black students obtained significantly lower mean ability estimates than White students.

This negative effect of KR appears to be contrary to the earlier reported data (Weiss, 1975) showing that KR itself eliminated mean racial differences in scores. What is similar between the two studies, however, is the finding that certain combinations of test administration conditions can reduce mean racial differences in ability estimates to nonsignificant levels. These results suggest that observed racial differences in verbal ability may be largely a function of test administration conditions, rather than a reflection of true racial differences.

The differences in the effects of KR on the Black students in this study and in the previous study may have been the result of differences in the way

KR was administered. In the earlier study the KR administered to both groups was designed to be specifically meaningful to the Black students. That is, KR was administered in terms which were derived from Black high school students, such as "right on." This form of feedback may have been more motivating to the Black students than the more typical feedback terms used in the present study. Black students in this study did report less favorable reactions to KR than White students, indicating that it "made them nervous" and "inhibited their concentration," thus potentially interfering with their test performance.

Another possible reason for the relatively high performance of Black students on the BR test under No-KR conditions and low performance under KR conditions relates to the item characteristics and difficulties of the tests. As mentioned above, the BR tests contained some words which were more appropriate for Black students, but the majority of the words represented only a reduction in the degree to which the items favored White students over the NBR test. Analysis of the nervousness reaction data indicated that the Black students were less nervous in the BR condition, presumably because some of the items appeared to be more appropriate for them. This effect was strongest for the paper-and-pencil test, as was the combined effect of bias-reduction and no knowledge of results for ability scores. While reduced nervousness may have aided performance on BR tests when No-KR was provided, BR performance was markedly reduced when KR was given, especially on the paper-and-pencil test. In the paper-and-pencil test this may have been due to the fact that the mean ability level for the Black students was lower than the ability level at which the conventional test was peaked. Thus, while the BR tests should have appeared to be easier for the Black students than the NBR tests, substantial negative feedback would have been received under the KR condition, possibly offsetting the positive psychological effects of taking the BR tests without receiving knowledge of results. When a Black student responded incorrectly to an item, in effect, the student was being told that he or she did not know the meaning of a "Black-type" word. It seems reasonable that negative feedback would have a stronger effect under these circumstances than in the NBR condition, an interpretation which is consistent with the result that Black students were less favorable to KR than White students.

This interpretation suggests that the motivational effects of KR may depend on the difficulty of the test for an examinee and, in particular, the proportion of negative versus positive feedback which the examinee receives. Figure 4 shows that for the Black students the negative effect of KR as provided in this study was stronger in the conventional paper-and-pencil test than in the computer-administered test. This may be due to the adaptive nature of the computer-administered test, which tends to equalize the amount of negative and positive KR each student receives, thus possibly reducing the adverse effects of negative KR.

The measurement properties of the BR tests were not as good as those of the NBR tests. Because of their item selection strategy, the NBR tests were substantially more discriminating than the BR tests. Related to this increased item discrimination was the increased precision of ability estimates in the NBR tests as indexed by the Bayesian posterior variances of these estimates. The lower levels of discrimination in the BR tests are consistent with the finding of Church et al. (1978) that "Black-type" words are less discriminating than more standard vocabulary test words for both Black and White students.

The data also permit some conclusions regarding conventional and adaptive testing strategies. Correlations of ability estimates across the two testing modes found substantial (r=.73), but not perfect, agreement between individual ability estimates. The distributions of the two sets of estimates suggested that divergence from stronger agreement was in part due to the adaptive test spreading individuals out more on the ability continuum than did the peaked tests. This may reflect the better measurement in the tails of the distribution, which is typical of adaptive tests. More equi-precise measurement was apparent in this study when the computerized adaptive and conventional paper-and-pencil tests were both non-bias-reduced. Under this condition, the ability estimates from the computer-administered adaptive test had smaller posterior variances except in the range of abilities where the paper-and-pencil test was peaked. For the BR tests, the paper-and-pencil test was more precise except for low-ability students. This differential effectiveness of the adaptive test under BR and NBR conditions implies that the selection of items within strata in stradaptive tests should be on the basis of item discriminations if the desired result is maximum precision, as has been suggested by Weiss (1974).

The data also show (Figure 4 and Table 10) that computerized adaptive testing also reduced the effects of other variables (e.g., KR, BR) on mean ability test performance in comparison to conventional paper-and-pencil test administration.

The clearest findings from the present study relate to the psychological effects of adaptive and conventional tests and the KR and BR variables on the two racial groups. The computer-administered adaptive test motivated both racial groups more than the conventional paper-and-pencil tests, as reflected in the significant main effect for the motivation dependent variable. The significant Mode×Order×Race×BR interaction for motivation scores (see Figure 13) indicated that under both BR and NBR conditions, the motivation level of the Black students was much lower on the paper-and-pencil test when it was taken second (Order 2). With the exception of the NBR adaptive test in Order 1, which was the only condition free of order effects, the Black students reported higher levels of motivation on the computerized adaptive test as compared to the paper-and-pencil test under both BR and NBR conditions. The strong order effect observed for the Black students was not generally found for the White students. For White students, motivation was highest for the adaptive test except when it was bias-reduced and was taken second.

The generally higher levels of reported motivation on the computer-administered adaptive test for both groups, but especially for the Black students, may have been a joint function of the novel testing format and the adaptive nature of the test; the test should have appeared less difficult than the conventional paper-and-pencil test, which was peaked above both racial groups' mean ability levels. The fact that the computerized adaptive test was able to actually increase motivation when it was given second, in contrast to the apparent fatigue effect (especially for Black students) when the paper-and-pencil test was given second, is especially encouraging for the use of this mode of test administration.

The data in Figure 13, and the marginally significant Mode×Order×Race effect which it subsumes, suggest that the motivation of Black students suffered more when the paper-and-pencil test was given second. The data also suggested that Black students preferred to take the paper-and-pencil test first and the computerized adaptive test second, while White students preferred the opposite. This significant Mode×Order×Race effect appeared elsewhere in the results. For the number of omitted responses variable, Black students omitted the most items when the paper-and-pencil test was taken second (see Figure 8) and the fewest items when the paper-and-pencil test was taken first. The opposite was true for the White students. Similarly for the guessing variable (see Figure 14), Black students reported guessing least (omitted more, were less motivated) when the paper-and-pencil test was administered second, while the White students guessed more when this test was administered first. These findings suggest that the differential sequential effect of the computer-administered and paper-and-pencil tests may be greater for the Black students. That is, once the novel computer-administered adaptive test had been given, the Black students seemed less interested in taking a conventional paper-and-pencil test. This would support the general conclusion of Johnson and Mihal (1973) that conditions of test administration can affect test-taking motivation.

Interestingly, while both Black and White students reported higher motivation on the computer-administered adaptive tests, they also reported more nervousness for this condition, as reflected in the significant main effect for the mode factor with the nervousness dependent variable. In fact, the significant Mode×Order interaction for nervousness (see Figure 10) indicated that when the computerized adaptive test was given first (Order 2), the increased nervousness carried over into the paper-and-pencil test, which was given second. When the paper-and-pencil test was given first (Order 1), nervousness was substantially lower until the computerized adaptive test was given, at which time it rose sharply. The higher reported motivation, but also nervousness, associated with administration of computerized adaptive tests suggests that during the administration of this test there was a general increased level of arousal or attention.

A further possible advantage of the computerized adaptive test over the conventional paper-and-pencil test was that students reported more guessing on the paper-and-pencil tests, which may be due to the fact that the adaptive test presented more items closer to the student's ability level. This apparent advantage resulted from the fact that the point at which the paper-and-pencil test had been peaked was above the ability level of the students. It is supported by the finding that higher ability students, besides reporting less nervousness, also reported less guessing.

A final interesting difference between the two modes of administration involves the differential relation of actual ability estimates to the various psychological reactions. Three of the four psychological dependent variables (reaction to knowledge of results, nervousness, and guessing) had statistically significant correlations with ability estimates in the expected direction. Thus, higher ability students reported more favorable reactions to knowledge of results, less nervousness, and less guessing. In all three cases, the relationship between estimated ability levels and psychological reactions was stronger for the conventional paper-and-pencil test. This supports the important conclusion that the computer-administered adaptive test was successful in reducing the effects of extraneous variables on test performance and is consistent with the findings and interpretation above, which suggested that Black students were less tolerant of paper-and-pencil tests and that both groups were more motivated on the adaptive test.

The data also showed racial differences in reactions to the provision of knowledge of results. While Black students felt less favorable about KR, as indicated earlier, a significant Race×KR interaction for the number of omitted responses score indicated that the presence of KR induced Black students to omit fewer items than under the No-KR condition. White students omitted the same average number of responses under both conditions. Thus, while KR made Black students more nervous, it also caused them to omit fewer responses. This implies that similar to the effects suggested above for computerized administration, the KR condition caused an increase in general arousal, or interest in one's performance. While this arousal could take the form of nervousness, reaction to KR was more favorable for both groups during the second test (see Figure 9), suggesting a familiarity effect.

Conclusions

Selection of items on the basis of an index of bias has been shown to reduce racial differences in mean performance on verbal ability tests when other variables, such as motivational factors, do not interfere with the effect. Since item selection based on bias-reduction alone can result in less precise measurement, simultaneous consideration of more traditional item statistics, such as item discrimination, should also be made in the development of bias-free tests.

The differential motivational impact of computer-administered versus conventional paper-and-pencil tests was given strong support in this study, and there were several indications that the psychological contrast between computer-administered and paper-and-pencil tests may differ for Black and White students. If this can be replicated, it may be possible to obtain more comparable motivational states across racial groups using computer-administered tests. In addition, the reaction to provision of knowledge of results differed for Black and White students.

This study has shown that ability test scores, and the reactions of different groups to ability tests, are to some extent a function of the conditions under which these tests are administered. The results support earlier studies on the effects of test administration conditions on both ability test scores and psychological reactions to testing (e.g., Betz & Weiss, 1976; Prestwood & Weiss, 1978). These data imply the need for further study of the effects of test administration conditions on members of minority groups to determine those administration conditions which maximize their ability estimates either directly or through their effects on the psychological environment of testing.

REFERENCES

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude (Research Bulletin RE 71-59). Princeton, NJ: Educational Testing Service, October 1971, pp. 1-24.

Betz, N. E. Prospects: New types of information and psychological implications. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. AD A027170).

Brown, J. M., & Weiss, D. J. An adaptive testing strategy for achievement test batteries (Research Report 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1977. (NTIS No. AD A046062).

Church, A. T., Pine, S. M., & Weiss, D. J. A comparison of levels and dimensions of performance in Black and White groups on tests of vocabulary, mathematics, and spatial ability (Research Report 78-3) Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978. (NTIS No. AD A062797).

DeWitt, L. J., & Weiss, D. J. A computer software system for adaptive ability measurement (Research Report 74-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD 773961).

Johnson, D. I., & Mihal, W. M. Performance of Blacks and Whites in computerized versus manual testing environments. American Psychologist, 1973, 28, 694-699.

Larkin, K. C., & Weiss, D. J. An empirical comparison of two-stage pyramidal adaptive ability testing (Research Report 75-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A027147).

Martin, J. T., Pine, S. M., & Weiss, D. J. An item bias investigation of a standardized aptitude test (Research Report 78-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement (Research Report 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD 781894).

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. AD A022964).

Owen, R. A Bayesian sequential procedure for quantal response in a context of adaptive verbal testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Pine, S. M., & Weiss, D. J. A comparison of the fairness of adaptive and conventional testing strategies (Research Report 78-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978. (NTIS No. AD A059436).

Prestwood, J. S., & Weiss, D. J. The effects of knowledge of results and test difficulty on ability test performance and psychological reactions to testing (Research Report 78-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-196.

Vale, C. D. Strategies of branching through an item pool. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD 768316).

Weiss, D. J. Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (NTIS No. AD A004270).

Weiss, D. J. (Ed.). Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975. (NTIS No. AD A018675).

APPENDIX: SUPPLEMENTARY TABLES

Table A
Examples of Vocabulary Items

Items from Black Literature and Black Psychologist			
Ranking		Gatemouth	
1.	Murdering	1.	Gossiper
2.	Exchange of insults	2.	Doorway
3.	Pig's intestines	3.	Jazz musician
4.	Fried cow's tail	4.	Dog
5.	Olympic event	5.	Fat person
Shiv		Swag	
1.	Politician	1.	Construction worker
2.	Genius	2.	Beggar
3.	Book	3.	Corrupt politician
4.	Drifter	4.	Stolen goods
5.	Knife	5.	Garbage
"White" type items from Webster's Seventh Collegiate Dictionary			
Borsch		Torte	
1.	Overcoat	1.	Cake
2.	Dog	2.	Twist
3.	Porter	3.	Shirt
4.	Soup	4.	Crime
5.	Chamber	5.	Answer
Afghan		Gefilte Fish	
1.	Alien	1.	Type of fish
2.	Harbor	2.	A game
3.	Canvas	3.	Food
4.	Vista	4.	A sport
5.	Blanket	5.	Sucker
Items from Standardized Vocabulary Tests			
Accumulate		Reinforce	
1.	Become cloudy	1.	Speak loudly
2.	Get angry	2.	Come again to
3.	Get dirty	3.	Revise
4.	Imitations	4.	Apply again
5.	Claws	5.	Make stronger
Oppressed		Capitulate	
1.	Wrinkled	1.	Entitle
2.	Expressed	2.	Surrender
3.	Musically talented	3.	Behead
4.	Disowned	4.	Put in charge
5.	Put down	5.	Congratulate

Table B

Item Numbers, Discrimination (*a*), Difficulty (*b*), and Bias Parameters for Items in the Vocabulary Stradaptive Item Pool (*c*=.20 for All Items)

Item	*a*	*b*	Bias
Stratum 7:			
(15 items)			
152	2.87	1.70	.35
162	1.58	1.57	-.18
166	1.19	2.07	.08
1256	1.12	1.52	.96
114	1.05	2.34	.06
1276	1.03	1.60	1.02
1229	1.01	2.40	-1.20
1220	.90	1.90	.26
1249	.64	2.31	1.53
1264	.58	2.54	.52
1206	.28	1.54	.09
1232	.26	3.18	-3.36
1245	.22	3.98	-4.44
1257	.22	1.61	-3.08
1278	.22	3.43	-3.97
Stratum 6:			
(17 items)			
1291	2.08	1.02	1.39
1303	1.52	1.14	1.07
43	1.49	1.00	1.57
1274	1.44	1.21	1.28
1243	1.22	1.23	1.28
1317	1.19	1.26	1.03
188	1.18	.95	1.01
1201	.98	1.01	.34
1240	.97	.91	.74
1222	.90	1.29	1.37
1217	.74	1.27	.88
1215	.72	1.26	1.35
1279	.68	1.05	1.61
1438	.64	1.36	.22
1309	.48	1.10	-.30
1305	1.33	1.14	1.03
52	1.70	1.46	.90
Stratum 5:			
(32 items)			
112	2.22	.32	.86
106	1.61	.51	.63
1430	1.27	.50	1.16
116	1.26	.88	.89
Stratum 5, *cont'd.*:			
1300	1.22	.71	.64
1227	1.19	.34	.90
1290	1.18	.34	.72
1425	1.17	.68	.85
1251	1.11	.84	2.00
1219	1.10	.38	.41
1244	1.06	.39	.51
127	1.05	.55	.52
1265	1.03	.32	1.55
1213	.98	.47	1.12
1321	.96	.68	.52
1325	.87	.69	1.09
1225	.84	.48	.40
1238	.79	.64	2.97
1297	.76	.45	1.04
1280	.75	.34	.16
95	.70	.40	1.40
648	.59	.87	.66
237	.48	.88	-.32
1204	.43	.52	.41
322	.41	.82	1.33
1202	.33	.65	-.10
1242	.09	.80	-4.57
1223	.09	.84	-5.46
1216	1.01	.57	.59
173	1.68	.33	.99
235	1.38	.86	1.03
51	1.76	.68	1.09
Stratum 4:			
(35 items)			
27	2.10	-.27	.59
123	2.03	.07	.25
1410	1.79	-.14	.98
1419	1.54	-.13	.75
1422	1.50	.01	1.04
190	1.46	-.24	.34
1426	1.36	-.09	.91
1207	1.34	.16	.44
47	1.32	.27	-.20
1420	1.32	.01	1.20
5	1.25	-.26	.80
1413	1.21	.25	1.11
Stratum 4, *cont'd.*:			
1315	1.16	-.05	.95
1429	1.11	.24	1.26
628	1.09	-.22	.76
1289	1.08	.17	1.39
1228	1.07	.26	.19
191	1.05	-.02	.24
1423	1.02	.01	.60
1298	1.01	-.28	.60
1284	1.01	-.09	.94
501	1.00	.03	.27
1260	.99	.24	-.12
1313	.89	-.23	.47
1268	.85	-.00	.11
1231	.85	.25	.77
1324	.84	-.10	.76
1236	.76	-.26	.51
1208	.74	.22	.92
63	.69	.11	.30
1316	.65	.25	1.46
1252	.56	-.11	-.54
1299	.53	-.03	-.20
199	.52	-.04	.54
101	2.07	.16	.62
Stratum 3.:			
(25 items)			
96	1.79	-.42	.74
23	1.42	-.45	1.18
1320	1.31	-.62	.60
1323	1.20	-.40	.68
86	1.16	-.36	.88
28	1.12	-.74	1.00
1293	1.07	-.47	.64
1414	1.04	-.73	.70
1403	1.04	-.89	1.09
32	.96	-.59	.94
1212	.95	-.87	1.33
1304	.94	-.31	.77
1246	.94	-.88	.62
1421	.93	-.42	.23
1295	.92	-.77	1.08
1310	.90	-.85	.34
16	.90	-.80	1.25
Stratum 3, *cont'd.*:			
1281	.84	-.38	1.40
1408	.82	-.31	1.11
1412	.81	-.60	1.01
202	.81	-.47	.88
1287	.73	-.47	1.05
88	.62	-.59	1.30
1248	.35	-.50	-1.11
1301	.28	-.53	-.21
Stratum 2:			
(16 items)			
1407	1.78	-1.42	.59
65	1.29	-1.06	.54
1409	1.25	-1.15	.92
1259	1.20	-.91	.30
182	1.14	-1.47	.70
25	1.12	-1.30	.96
1405	1.09	-1.04	.59
1282	1.08	-.91	.79
1235	.79	-.94	.01
1311	.77	-1.22	1.35
1262	.76	-.94	-.35
1322	.74	-.96	1.02
19	.73	-1.20	.73
1312	.62	-.91	.82
1318	.60	-1.05	1.06
1406	.53	-.92	.08
Stratum 1:			
(11 items)			
1400	1.49	-1.64	.26
1404	1.32	-1.57	.44
1402	1.13	-1.79	.58
122	1.10	-1.91	.77
1234	.90	-1.83	.66
71	.75	-1.94	2.05
66	.69	-1.64	-.20
1272	.63	-1.59	.30
1275	.56	-2.04	-.03
1239	.52	-1.66	-.36
240	.41	-1.95	3.73

Note. For the bias-reduced adaptive test, the items were ordered within each stratum in increasing order of bias.

Table C
Item Numbers, Discrimination (*a*), Difficulty (*b*), and Bias Parameters for Items in the Vocabulary Pencil-and-Paper Tests

	Non-Bias-Reduced				Bias-Reduced		
Item No.	*a*	*b*	Bias	Item No.	*a*	*b*	Bias
52	1.70	1.46	.90	1204	.43	.52	.41
1302	1.50	.87	1.21	1414	1.04	.73	.70
1418	1.52	.13	.84	501	1.00	.03	.27
189	1.78	.60	1.71	1211	2.27	.18	.28
85	1.54	.29	.39	1223	.09	.84	-5.46
22	1.42	- .13	.90	1260	.99	.24	- .12
311	1.50	.12	.22	47	1.32	.27	- .20
1305	1.33	1.14	1.03	1240	.97	.91	.74
105	1.37	- .40	1.12	181	1.17	- .66	.65
1401	1.59	-1.48	.37	1401	1.59	-1.48	.37
1411	1.85	.09	1.24	191	1.05	- .02	.24
1254	1.68	- .60	.95	1323	1.20	- .40	.68
285	1.48	.35	.74	1201	.98	1.01	.34
1415	1.29	- .69	.76	1219	1.10	.38	.41
1416	2.04	- .21	.74	1228	1.07	.26	.19
1211	2.27	.18	.28	1244	1.06	.39	.51
51	1.76	.68	1.09	311	1.50	.12	.22
522	1.36	.12	.62	1299	.53	- .03	- .20
121	1.18	- .92	.89	1235	.79	- .94	.01
181	1.17	- .66	.65	1248	.35	- .50	-1.11

Table D
Test Reaction Questions, by Scale

Scaled Score	Knowledge of Results Scale
	DID RECEIVING FEEDBACK AFTER EACH QUESTION INTERFERE WITH YOUR ABILITY TO CONCENTRATE ON THE TEST?
1	□ NO, NOT AT ALL
2	□ YES, SOMEWHAT
3	□ YES, MODERATELY SO
4	□ YES, VERY MUCH SO
	DID GETTING FEEDBACK AFTER EACH QUESTION MAKE YOU NERVOUS?
1	□ NO, NOT AT ALL
2	□ YES, SOMEWHAT
3	□ YES, MODERATELY SO
4	□ YES, VERY MUCH SO

Scaled Score	Nervousness Scale
	WERE YOU NERVOUS WHILE TAKING THE TEST?
1	□ NOT AT ALL
2	□ SOMEWHAT
3	□ MODERATELY SO
4	□ VERY MUCH SO
	DID NERVOUSNESS WHILE TAKING THE TEST PREVENT YOU FROM DOING YOUR BEST?
4	□ YES, DEFINITELY
3	□ YES, SOMEWHAT
2	□ PROBABLY NOT
1	□ DEFINITELY NOT

Scaled Score	Motivation Scale
	DID YOU CARE HOW WELL YOU DID ON THE TEST?
4	□ I CARED A LOT
3.2	□ I CARED SOME
2.4	□ I CARED A LITTLE
1.6	□ I CARED VERY LITTLE
.8	□ I DIDN'T CARE AT ALL

Scaled Score	Motivation Scale (cont'd.)
	DID YOU FEEL CHALLENGED TO DO AS WELL AS YOU COULD ON THE TEST?
1	□ NOT AT ALL
2	□ SOMEWHAT
3	□ FAIRLY MUCH SO
4	□ VERY MUCH SO
	WERE YOU INTERESTED IN KNOWING WHETHER YOUR ANSWERS WERE RIGHT OR WRONG?
4	□ I WAS VERY INTERESTED
3	□ I WAS MODERATELY INTERESTED
2	□ I WAS SOMEWHAT INTERESTED
1	□ I DIDN'T CARE AT ALL

Scaled Score	Guessing Scale
	ON HOW MANY OF THE QUESTIONS DID YOU GUESS?
4	□ ALMOST ALL OF THE QUESTIONS
3.33	□ MORE THAN HALF OF THE QUESTIONS
2.67	□ ABOUT HALF OF THE QUESTIONS
2	□ LESS THAN HALF OF THE QUESTIONS
1.33	□ ALMOST NONE OF THE QUESTIONS
.67	□ NONE OF THE QUESTIONS
	HOW OFTEN WERE YOU SURE THAT YOUR ANSWERS TO THE QUESTIONS WERE CORRECT?
.8	□ 1. ALMOST ALWAYS
1.6	□ 2. MORE THAN HALF OF THE TIME
2.4	□ 3. ABOUT HALF OF THE TIME
3.2	□ 4. LESS THAN HALF OF THE TIME
4	□ 5. ALMOST NEVER

Table E
Means and Standard Deviations of Bayesian Posterior Ability Estimates as a Function of Ability Estimates for Adaptive and Conventional Tests in Non-Bias-Reduced and Bias-Reduced Conditions

Bayesian Ability Estimate Interval		Non-Bias-Reduced Condition						Bias-Reduced Condition					
		Conventional Test			Adaptive Test			Conventional Test			Adaptive Test		
Lo	Hi	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
-2.0	-1.81	0	--	--	7	.080	.006	2	.169	.002	6	.114	.016
-1.8	-1.61	2	.103	.000	2	.089	.012	8	.195	.017	4	.131	.012
-1.6	-1.41	7	.116	.013	6	.092	.026	12	.180	.035	6	.122	.028
-1.4	-1.21	19	.114	.016	8	.078	.008	10	.158	.037	6	.142	.033
-1.2	-1.01	14	.115	.018	9	.077	.009	10	.189	.031	9	.134	.013
-1.0	- .81	20	.106	.025	13	.092	.024	13	.174	.043	8	.177	.054
- .8	- .61	12	.098	.023	16	.082	.008	11	.143	.018	6	.152	.033
- .6	- .41	8	.089	.021	15	.103	.034	4	.130	.012	9	.160	.031
- .4	- .21	8	.088	.015	9	.106	.037	12	.150	.014	15	.180	.027
- .2	- .01	11	.073	.046	7	.093	.007	15	.114	.062	10	.136	.100
.0	.19	4	.086	.008	10	.097	.028	3	.132	.016	8	.184	.029
.2	.39	4	.079	.003	3	.082	.017	4	.128	.005	7	.227	.054
.4	.59	2	.087	.008	1	.068	.000	1	.140	.000	5	.233	.024
.6	.79	0	--	--	0	--	--	3	.139	.008	0	--	--
.8	.99	0	--	--	0	--	--	2	.155	.005	2	.289	.028
1.0	1.19	2	.139	.009	0	--	--	0	--	--	0	--	--
1.2	1.39	0	--	--	1	.134	.000	0	--	--	0	--	--
1.4	1.59	0	--	--	0	--	--	0	--	--	0	--	--
1.6	1.79	0	--	--	0	--	--	0	--	--	0	--	--
1.8	2.00	0	--	--	1	.209	.000	0	--	--	1	.204	.000

Table F
Means and Standard Deviations of Bayesian Ability Estimates
for all Combinations of the Independent Variables

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	-.94	-.86	-.59	-.90	-.88	-.81	-1.13	-.82
S.D.	1.02	.95	1.07	.93	.84	.94	.57	.58
P&P								
N	15	14	13	14	15	13	12	14
Mean	-1.05	-.98	-.51	-.73	-.79	-.99	-.91	-.84
S.D.	.66	.77	.81	.62	.48	.71	.50	.62
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	-.75	-.32	-.77	-.56	-.58	-.60	-.72	-.63
S.D.	.67	.78	.66	.88	.64	.74	.75	.70
P&P								
N	13	14	11	13	14	15	15	13
Mean	-.66	-.40	-1.04	-.58	-.78	-.68	-.42	-.55
S.D.	.68	.54	.51	.62	.45	.50	.67	.57

Table G
Means and Standard Deviations of Bayesian Posterior Variances
for all Combinations of the Independent Variables

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	.15	.17	.16	.16	.09	.11	.09	.08
S.D.	.04	.04	.04	.04	.02	.04	.01	.02
P&P								
N	15	14	13	14	15	13	12	14
Mean	.18	.16	.16	.15	.10	.11	.10	.10
S.D.	.04	.04	.03	.03	.03	.02	.02	.01
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	.15	.19	.15	.20	.11	.09	.09	.08
S.D.	.03	.06	.03	.06	.04	.02	.02	.01
P&P								
N	13	14	11	13	14	15	15	13
Mean	.17	.14	.17	.15	.11	.09	.10	.10
S.D.	.04	.03	.03	.04	.02	.02	.03	.03

Table H
Means and Standard Deviations of Number of Omitted Responses
Under All Combinations of the Independent Variables

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	15	13	13	13	15	13	12	14
Mean	3.00	1.38	3.08	3.38	2.13	1.23	4.33	4.50
S.D.	3.57	2.02	2.93	3.50	2.62	1.92	4.10	2.98
P&P								
N	15	14	13	14	15	13	12	14
Mean	2.07	2.86	1.77	4.50	1.87	.23	4.42	5.00
S.D.	4.50	4.79	2.89	6.21	2.82	.60	5.45	6.67
Whites								
CAT								
N	12	15	11	14	14	15	15	15
Mean	3.17	2.27	2.73	3.00	2.21	1.73	3.33	3.60
S.D.	4.49	2.28	2.80	2.42	2.72	1.98	4.62	3.18
P&P								
N	13	14	11	13	14	15	15	13
Mean	3.46	1.93	2.82	2.46	5.21	2.27	2.53	2.85
S.D.	5.08	2.76	3.84	3.50	6.23	4.25	4.27	4.26

Table I
Means and Standard Deviations of Dependent Variables for the Combined Racial, Bias Reduction, Knowledge of Results, Order of Administration, and Mode of Administration Groups

	Bayesian Scores			Posterior Variance			Number of Omits		
Combined Groups	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
Racial									
Blacks									
CAT	108	-.87	.86	108	.12	.05	108	2.87	3.14
P&P	110	-.85	.66	110	.13	.04	110	2.83	4.75
Whites									
CAT	112	-.61	.72	112	.13	.06	112	2.73	3.12
P&P	108	-.63	.58	108	.13	.04	108	2.94	4.36
Bias Reduction									
Bias-Reduced									
CAT	107	-.70	.88	107	.17	.05	107	2.72	3.03
P&P	107	-.74	.68	107	.16	.03	107	2.73	4.31
Non-Bias-Reduced									
CAT	113	-.76	.73	113	.09	.03	113	2.87	3.24
P&P	111	-.74	.58	111	.10	.02	111	3.03	4.80
Knowledge of Results									
Knowledge of Results									
CAT	112	-.71	.83	112	.13	.05	112	2.14	2.78
P&P	113	-.79	.62	113	.13	.04	113	2.49	4.29
No Knowledge of Results									
CAT	108	-.76	.78	108	.13	.05	108	3.47	3.34
P&P	105	-.69	.64	105	.13	.04	105	3.30	4.81
Order of Administration									
P&P/CAT									
CAT	107	-.79	.80	107	.12	.04	107	2.97	3.50
P&P	108	-.76	.63	108	.14	.04	108	2.98	4.52
CAT/P&P									
CAT	112	-.68	.81	112	.14	.06	112	2.65	2.74
P&P	110	-.72	.63	110	.13	.04	110	2.78	4.60
Mode of Administration									
CAT	220	-.73	.80	220	.13	.05	221	2.78	3.12
P&P	218	-.74	.63	218	.13	.04	218	2.88	4.55

Table J
Means and Standard Deviations of
Knowledge of Results Scale Scores

	Bias-Reduced		Non-Bias-Reduced	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks				
CAT				
N	14	13	13	12
Mean	1.89	1.73	1.81	2.08
S.D.	.90	.90	.80	1.18
P&P				
N	13	14	15	13
Mean	2.31	1.64	1.83	1.69
S.D.	1.03	.72	.79	.75
Whites				
CAT				
N	12	13	12	14
Mean	1.38	1.58	1.38	1.54
S.D.	.43	.61	.53	.54
P&P				
N	11	14	13	12
Mean	1.46	1.29	1.31	1.46
S.D.	.47	.47	.44	.54

Table K
Means and Standard Deviations of the Nervousness Scale Scores

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	2.46	1.85	2.00	1.75	1.85	2.29	2.25	1.93
S.D.	.91	.69	.54	.51	.75	.72	.72	.80
P&P								
N	15	14	13	14	15	13	12	14
Mean	1.93	1.89	1.73	1.57	1.90	2.27	2.21	1.96
S.D.	.94	1.08	.70	.76	.71	.75	.78	.80
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.08	2.00	2.18	1.83	1.96	1.93	1.71	2.23
S.D.	.63	.68	.46	.62	.66	.62	.47	.56
P&P								
N	12	14	11	13	14	14	13	13
Mean	1.88	1.96	1.86	1.73	1.75	2.14	1.58	2.15
S.D.	.64	.66	.50	.60	.55	.82	.70	.56

Table L
Means and Standard Deviations of the Motivation Scale Scores

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	3.52	2.66	2.90	3.14	3.02	3.27	3.17	2.83
S.D.	.56	.92	.49	.81	.47	.45	.48	.74
P&P								
N	15	14	13	14	15	13	12	14
Mean	3.37	2.34	2.90	3.17	2.96	2.62	3.02	2.66
S.D.	.64	.88	.61	.81	.66	.52	.60	.70
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.62	3.07	2.63	2.98	2.87	2.86	3.12	3.15
S.D.	.77	.71	.51	.56	.88	.82	.94	.59
P&P								
N	12	14	11	13	14	12	13	13
Mean	2.77	3.01	2.92	2.75	2.68	2.69	2.78	3.03
S.D.	.98	.52	.83	.56	.91	.76	.95	.48

Table M
Means and Standard Deviations of the Guessing Scale Scores

	Bias-Reduced				Non-Bias-Reduced			
	Knowledge of Results		No Knowledge of Results		Knowledge of Results		No Knowledge of Results	
Group and Mode	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P	Order 1 P&P/CAT	Order 2 CAT/P&P
Blacks								
CAT								
N	14	13	13	14	13	12	12	14
Mean	2.44	2.08	2.15	1.87	2.10	2.24	2.24	1.97
S.D.	.78	.52	.59	.86	.76	.54	.70	.67
P&P								
N	15	14	13	14	15	13	12	14
Mean	2.32	2.37	2.07	2.33	2.20	2.58	2.77	2.12
S.D.	.86	.71	.66	.96	.80	.72	.58	.69
Whites								
CAT								
N	12	13	11	12	12	14	14	13
Mean	2.35	2.12	2.27	2.48	2.07	2.30	2.29	2.37
S.D.	.66	.72	.69	.49	.53	.62	1.02	.69
P&P								
N	12	14	11	13	14	12	13	13
Mean	2.30	2.19	2.27	2.48	2.48	2.18	2.63	2.27
S.D.	.93	.53	.97	.53	.79	.87	.87	.48

Table N

Means and Standard Deviations of the Test Reaction Scale Scores for the Combined Racial, Bias Reduction, Knowledge of Results, Order of Administration, and Mode of Administration Groups

	Knowledge of Results			Nervousness			Motivation			Guessing		
Combined Groups	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*	*N*	Mean	*S.D.*
Racial												
Blacks												
CAT	52	1.88	.93	105	2.04	.73	105	3.06	.67	105	2.14	.69
P&P	57	1.85	.84	110	1.93	.83	110	2.88	.74	110	2.33	.77
Whites												
CAT	51	1.47	.52	101	1.98	.59	101	2.92	.74	101	2.28	.69
P&P	50	1.37	.47	104	1.88	.65	102	2.83	.75	102	2.35	.77
Bias Reduction												
Bias-Reduced												
CAT	52	1.65	.75	102	2.02	.66	102	2.96	.72	102	2.22	.68
P&P	54	1.67	.79	106	1.82	.75	106	2.91	.77	106	2.29	.78
Non-Bias-Reduced												
CAT	51	1.70	.82	104	2.01	.68	104	3.03	.70	104	2.20	.70
P&P	53	1.58	.67	108	1.99	.73	106	2.81	.71	106	2.40	.75
Knowledge of Results												
Knowledge of Results												
CAT	103	1.68	.78	103	2.05	.72	103	2.99	.75	103	2.22	.64
P&P	105	1.63	.73	111	1.96	.78	109	2.82	.78	109	2.33	.79
No Knowledge of Results												
CAT				103	1.98	.61	103	3.00	.67	103	2.20	.74
P&P				103	1.84	.70	103	2.90	.70	103	2.36	.75
Order of Administration												
P&P/CAT												
CAT	51	1.63	.73	101	2.06	.68	101	3.00	.70	101	2.24	.72
P&P	53	1.74	.81	105	1.85	.71	105	2.93	.78	105	2.38	.82
CAT/P&P												
CAT	52	1.72	.84	105	1.97	.66	105	2.99	.72	105	2.17	.66
P&P	54	1.51	.63	109	1.96	.78	107	2.79	.70	107	2.31	.72
Mode of Administration												
CAT	103	1.68	.78	206	2.02	.67	206	2.99	.71	206	2.21	.69
P&P	107	1.63	.73	214	1.91	.74	212	2.86	.74	212	2.34	.77