DOCUMENT RESUME

ED 129 864 TM 005 638

AUTHOR	Betz, Nancy E.; Weiss, David J.
TITLE	Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. Research Report 76-3.
INSTITUTION	Minnesota Univ., Minneapolis. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE	Jun 76
CONTRACT	N00014-76-C-0243; NR150-382
NOTE	47p.; For a related document, see TM 005 637
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Ability; Achievement Tests; Branching; College Students; *Computer Oriented Programs; *Feedback; *Performance Factors; Response Style (Tests); Statistical Analysis; *Testing; Verbal Ability
IDENTIFIERS	*Computer Assisted Testing

ABSTRACT

The effects of immediate knowledge of results (KR) concerning the correctness or incorrectness of each item response on a computer-administered test of verbal ability were investigated. The effects of KR were examined on a 50-item conventional test and a stradaptive ability test and in high- and low-ability groups. The primary dependent variable was maximum likelihood ability estimates derived from the item responses. Results indicated that mean test scores for the High-Ability group receiving KR were higher than for the No-KR group on both the conventional and stradaptive tests. For Low-Ability examinees, mean scores were higher under KR conditions than under NO-KR conditions on both tests, but the difference was statistically significant only for the conventional test. However, the higher mean scores of the Low-Ability testees on the stradaptive test indicated that for low-ability examinees, adaptive testing had the same effects on test performance as did the provision of immediate KR. The results of the study were interpreted as indicating the potential of both immediate knowledge of results and adaptive testing procedures to increase the extent to which ability tests measure "maximum performance" levels. (Author)

EFFECTS OF IMMEDIATE KNOWLEDGE OF RESULTS AND ADAPTIVE TESTING ON ABILITY TEST PERFORMANCE

Nancy E. Betz
and
David J. Weiss

RESEARCH REPORT 76-3
JUNE 1976

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

Prepared under contract No. N00014-76-C-0243, NR150-382
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
Research Report 76-3

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Nancy E. Betz and David J. Weiss

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-0243

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.:NR150-382

11. CONTROLLING OFFICE NAME AND ADDRESS
Personnel and Training Research Programs
Office of Naval Research
Arlington, Virginia 22217

12. REPORT DATE
June 1976

13. NUMBER OF PAGES
36

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

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)

testing	sequential testing	programmed testing
ability testing	branched testing	response-contingent testing
computerized testing	individualized testing	automated testing
adaptive testing	tailored testing	

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

This study investigated the effects of immediate knowledge of results (KR) concerning the correctness or incorrectness of each item response on a computer-administered test of verbal ability. The effects of KR were examined on a 50-item conventional test and a stradaptive ability test and in high- and low-ability groups. The primary dependent variable was maximum likelihood ability estimates derived from the item responses. Results indicated that mean test scores for the High-Ability group receiving KR were higher than for

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the No-KR group on both the conventional and stradaptive tests. For Low-Ability examinees, mean scores were higher under KR conditions than under No-KR conditions on both tests, but the difference was statistically significant only for the conventional test. However, the higher mean scores of the Low-Ability testees on the stradaptive test indicated that for low-ability examinees, adaptive testing had the same effects on test performance as did the provision of immediate KR. Knowledge of results did not have significant effects on either response latencies, response consistency on the stradaptive test, or the internal consistency reliability of the conventional test. No significant score difference were found on a 44-item post-test administered without KR, indicating that the facilitative effects of knowledge of results on test performance were confined to the test in which KR was provided. The results of the study were interpreted as indicating the potential of both immediate knowledge of results and adaptive testing procedures to increase the extent to which ability tests measure "maximum performance" levels.

CONTENTS

Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance

The description of ability tests as measures of "maximum performance" (Cronbach, 1970) implies that such tests should reflect the highest level of performance of which a given individual is capable. According to Cronbach, the distinguishing feature of such tests "is that the subject is encouraged to earn the best score he can," (p. 35). Thus, to the extent that individuals do not perform to their fullest capabilities on an ability test, the measurement of those individuals' ability levels may be less accurate, and the predictive validity of the obtained scores may be reduced.

It is reasonable to assume that examinees will perform to their fullest capabilities on an ability test only if they are motivated to do so. According to Samuda (1975):

> A person who is being tested usually tries to do his best. Therefore, motivation is one of the *a priori* assumptions upon which tests are built. The great majority of available data leads to the observation that motivation has a determining effect upon level of performance. Thus differences in performance may be attributed, in part, to differences in motivation. (p. 82)

The importance of maintaining examinee motivation at high levels was recognized in the early days of ability testing by the constructors of individually administered intelligence tests. Terman (1916) recommended that frequent praise was essential to maintain high levels of motivation in the administration of the Stanford-Binet test. Recent versions of the manuals for the Stanford-Binet (Terman & Merrill, 1960) and the WAIS (Wechsler, 1955) instruct the examiner to give frequent praise and encouragement to the examinee. Thus, means of maintaining high examinee motivation have always been perceived as an important aspect of the administration of individual intelligence tests. However, such tests have also been characterized by wide differences among examiners in both administration and scoring (see Sattler & Theye, 1967, and Weiss & Betz, 1973, for reviews of this literature).

By the end of World War I, group testing had become the predominant means of measuring intelligence and abilities (DuBois, 1970). Group tests, while characterized by a very high degree of standardization and objectivity, had no provision for maintaining high levels of testee effort and motivation.

The provision of immediate knowledge of results (KR) is one means of possibly increasing motivation that can be incorporated into group-administered paper-and-pencil tests. Immediate KR has a long history in the study of human learning and performance. In fact, the facilitative effect of KR in human learning is one of the best established findings in the research literature (e.g., Ammons, 1956; Annett, 1961; 1969; Bilodeau & Bilodeau, 1961).

Knowledge of results on classroom tests has been hypothesized to be important in motivating classroom achievement (e.g., Ross, 1933) and in facilitating learning and retention of learned material. Many studies have been concerned with the effects of various delays in returning test results to students on their subsequent classroom performance and achievement (e.g., Brown, 1932; Kulhavy & Anderson, 1972; McMahon, 1973; Newman, Williams, & Hiller, 1974; Plowman & Stroud, 1942). Ammons (1956) and Annett (1969) have reviewed the literature on the effects of KR in experimental studies of learning while Annett (1969) and Geis and Chapman (1971) reviewed research relevant to the use of KR in programmed instruction.

However, little published research has dealt with the effects of KR on performance on tests of intelligence or ability. Given the importance ascribed to KR by researchers studying other aspects of human performance, it is surprising that so little attention has been directed toward the possible effects of immediate, on-going KR on individuals' demonstrations of their fullest intellectual capabilities, in the situation in which those capabilities are being assessed.

Knowledge of Results in Group-Administered Tests

Methods of providing KR. The earliest devices used to provide KR on objective multiple-choice or true-false tests were developed beginning in 1915 by Pressey (1926). Pressey's interest, however, was in KR as a teaching device rather than as a means for motivating high levels of test performance.

These early devices developed by Pressey, called "mechanical instructors", soon were replaced by the Pressey (1950) punchboard. This device consisted of a top punchboard, slotted with as many holes as there were alternative answers to the questions, and a bottom punchboard with holes only for the correct answers. When the examinee punched a hole for the correct answer, an answer sheet lying between the two boards was perforated. As with the mechanical instructor, the examinee was required to select alternatives until the correct one was found.

Pressey (1950) described a series of studies concerning the effectiveness of the punchboard in facilitating learning. These experiments, which contrasted the performance of examinees using punchboards with that of examinees using standard answer sheets, indicated that the punchboards facilitated learning in terms of such criteria as direct and free recall of tested material. Pressey also found that students liked using the punchboards, came to depend on the immediate appraisals, and became frustrated when the later use of standard tests left them without knowledge of results.

More recent examples of KR devices include Montor's (1970) "Trainer-Tester", and Lord's (1971) "flexilevel" test. The Trainer-Tester uses answer sheets on which the testees erase the ink covering their response choices and thereby are informed immediately of the correctness of their responses. The Trainer-Tester requires the testee to continue to erase answers until the correct one is chosen, thereby facilitating learning.

Lord's flexilevel test, on the other hand, was not originally intended as a KR device. Instead, it was an implementation of an adaptive or tailored test in a paper-and-pencil, rather than computer-administered, format. The flexilevel test utilizes a specially constructed answer sheet to facilitate the adaptive item-administration procedure. When examinees choose wrong answers, a red spot appears on the answer sheet. When a correct answer is chosen, a blue spot appears. The color of the spot directs the examinee to the next test item to be answered.

While this format does not provide direct KR, testees are likely to realize early in the test that items that were easy for them were followed by the blue (correct) spot, while items that were difficult were followed by the red (incorrect) spot. This knowledge can then be generalized to item responses about which they are unsure. To date there has been only one study in which the flexilevel test was actually administered by paper and pencil (Olivier, 1974), but the study was not concerned with the possible effects of the immediate KR that the examinees were probably receiving.

Thus, while methods for implementing KR during ability testing have long been available, most studies have been concerned with its effects on learning and retention. Few studies have examined the function of KR as an incentive enhancing the immediate performance of individuals on objective tests. The studies that do exist utilized classroom achievement tests rather than intelligence or ability tests.

Effects of KR in achievement tests. A study by Bierbaum (1965) utilized a Pressey-type punchboard to study the effects of immediate KR on test performance. Two parallel classroom tests were administered to a class of 23 students. On the first test, half the students received KR and the other half did not. On the second test, this condition was reversed so that those who initially had received KR did not, and vice versa. Results indicated that significantly more errors were made on the KR items than on the no-KR items, and this finding was similar in degree for both KR-first and KR-second groups. Further investigation revealed that students considered the KR condition to "put them under more pressure", and Bierbaum concluded that KR may cause increased anxiety. However, in this study students had to continue to select answers until they found the correct one; it is possible that this requirement enhanced the pressure they felt to choose the correct answer.

Heald (1970) studied the effects of KR on achievement test performance and upon retention of learned material as measured by a retest after one week. In Heald's study, two different KR conditions were contrasted with a control condition. In the "KR-Reference" condition, examinees were informed of the correctness of their responses. If an answer was incorrect they were referred immediately to the passage in the class text which addressed that item; following reference to the text, they were to respond a second time to the item. In the "KR-Alone" condition, examinees were informed whether or not they were correct and were required to continue responding until they answered the item correctly. The control condition utilized a standard answer sheet format. Fifty-four students in a graduate level course in educational administration were tested on material relevant to the topic "audiovisual materials for teachers." Students were classified into high and low test anxiety groups on the basis of the Sarason Text Anxiety Scale and were assigned randomly to treatment conditions from within each anxiety group.

Results indicated that KR had significant effects on performance in both initial test and retest. The KR-Reference condition led to higher test scores than did the KR-Alone condition, and both KR groups performed better than did students in the control condition on both tests. There were no significant differences in performance as a function of anxiety level, nor was there any significant interaction between KR condition and anxiety level.

In Beeson's (1973) study, students were administered tests in which one half of the items were followed by immediate KR and the other half were given delayed, post-test, KR. Immediate KR was administered using an IBM card punchboard. Three groups of students, two college groups and one junior high school group, were studied, using mathematics achievement tests. Within each group, half the students received immediate KR on the first half of the items and delayed KR on the second half; the other half of the group received the KR conditions in reverse order. This procedure was followed on each of 10 one-hour exams and a final exam. The order of KR for each subgroup was counterbalanced so that no subject received immediate KR on the same half of the test in any two consecutive tests.

Results indicated that there were no significant differences within any of the 10 one-hour exams, but that the performance on the immediate KR half of the test was significantly better ($p<.05$) on the final exam. In general, performance was better when students were given immediate KR, and Beeson attributes the significance of the difference found only on the final exam to the fact that it was a longer and more reliable test.

Spencer and Barker (1969) studied the effects of immediate KR on retention of learned material over a time interval. While amount retained was the major dependent variable of interest, this study also provides data relevant to the effects of KR on test performance itself. On the first test given (an achievement test in biology), one group received item-by-item feedback using a punchboard answer sheet, while the other group used a regular answer sheet. On the retest given 18 days later, all students used the regular answer sheet. It was found that the group using the punchboard scored significantly lower on the initial test than did the control group. But on the retest the experimental group scored significantly higher than did the control group.

One major problem in the study of KR in teacher-constructed tests has been the failure to control for the possibility that the KR received on one test item may provide the examinee with information concerning the correct answers to succeeding items. For example, in Heald's (1970) study reporting the facilitative effects of KR, the number of relevant cues provided by the KR easily could have accounted for its beneficial effects. Similarly, the retest difference in Spencer and Barber's study could have resulted from the learning effects of KR.

One study explicitly designed to separate these two effects was that by Strang and Rust (1973), who studied the effects of KR on an achievement test of course-related facts and their applications. The items were constructed so that knowledge of results on one item would not provide clues to the answers of succeeding questions. Thus, the interest was solely in KR as a motivational variable. In this study, both experimental and control groups were first administered a 25-item test under no-KR conditions. The students then were divided into four groups, resulting from a cross-classification of task definition (test vs. experimental exercise) and knowledge of results vs. no-KR. In the two KR groups, students indicated their answers by erasing one of five answer spots. If the answer was correct, a "+" appeared, and if it was incorrect, the letter corresponding to the correct choice appeared. The results of the 2 x 2 analysis of covariance of the scores on the second 25-item test (using scores on the first 25-item test as a covariate) indicated that students in the KR condition made significantly more errors. Additionally, students in the KR condition reported significantly more nervousness during testing than did students in the no-KR condition. Strang and Rust hypothesized that the increase in errors under KR may have been caused by the greater nervousness of the examinees.

Knowledge of Results in Individually Administered Tests

In addition to the studies of KR on classroom achievement tests, two studies have introduced KR into the administration of individual intelligence tests. Sweet and Ringness (1971) administered the Wechsler Intelligence Scale for Children (WISC) to elementary school boys under one of three conditions. In the first condition, the WISC was administered in the "standard" manner. In the second condition, students were told by the examiner if a response had been "correct" or "mostly correct"; the examiner made no response when an answer was incorrect. The third condition utilized the award of a poker chip, exchangeable later for money, following each correct response.

Results indicated that there were no differential treatment effects for middle-class whites or for lower-class blacks, but that lower-class whites performed significantly better when reinforced, either with KR or with poker chips, for their correct responses. Sweet and Ringness explained their results by concluding, first, that middle-class children already perform at a high level under standard administrative conditions and do not profit from the additional motivation provided by incentive conditions. Second, the lack of a treatment effect for the lower-class black group may have been due to the fact that all examiners were white females. Literature reviewed by Sattler and Theye (1967) and Weiss and Betz (1973) has shown that performance on intelligence tests can be affected by the race of the examiner and/or by interactions between examiner and examinee race. Furthermore, these same effects and interactions have been found for examiner/examinee sex factors, and in the Sweet and Ringness study all students were male and all examiners were female.

A study employing greater standardization of administrative procedures was reported by Zontine, Richards, and Strang (1972). In their study, all instructions and test items for the Peabody Picture Vocabulary Test (PPVT) were presented by tape recorder to a group of 72 seven- to eight-year-old children. The role of the examiners was limited to the recording of answers, regulating the speed of the tape recorder, and in the experimental conditions controlling the administration of the reinforcer.

All 72 children were administered Form A of the PPVT by tape recorder without reinforcement. Following the administration of Form A, which served as control and covariate in the data analysis, the children were assigned randomly to one of three conditions for the administration of Form B of the PPVT two months later. Examinees in Group 1 received Form B under conditions identical to those of the administration of Form A. Students in Group 2 were given immediate KR in the form of a white light following each correct response; after each five white lights, a red light was turned on to indicate to the examinees their cumulative levels of performance. Test administration to students in Group 3 was the same as that of Group 2, but in addition these students were given a food reward after earning each red light; thus this condition added an extrinsic reward to the KR given. Analysis of variance of the difference scores between Form A and Form B and analysis of covariance of the Form B scores using Form A scores as the covariate showed no significant differences in Form B performance as a function of differential treatments.

Knowledge of Results in Computer-Administered Objective Tests

With the advent of interactive computer systems has come the capability of administering tests by computer. One important potential advantage of computer-assisted testing procedures in the area of ability measurement is the ease with which examinees can be provided immediate information about whether their responses to each test item were correct or incorrect. Bayroff (1964), Ferguson and Hsu (1971), and Weiss and Betz (1973) have suggested that the provision of immediate KR may have positive motivating effects on examinees.

In spite of the ease with which immediate KR can be provided during the administration of an ability test by an interactive computer, only one study has investigated the effects of providing KR on a computer-administered test. In this study[1] (Betz, 1975), a group of 90 inner-city high school students, consisting of 27 black and 53 white students, were administered two vocabulary tests by computer. One test consisted of 40 items that were generally somewhat too difficult for the average testee. The other test administered was a 15 item "pyramidal" (Weiss, 1974, pp. 12-17) adaptive test. The manipulated independent variables in this study were: 1) whether or not immediate KR was given and 2) whether the conventional 40-item test or the 15-item pyramidal test was administered first. The group was classified by race into black and white sub-groups.

The results of a three-way (2 x 2 x 2) analysis of variance on the conventional test scores showed a main effect only for race; the level of performance of whites was significantly higher than that of blacks. None of the two-way interactions was significant, but there was a significant three-way interaction between race, order, and feedback. Analysis of the sub-group means indicated that under KR conditions when the conventional test had been administered first, the mean score obtained by the blacks (26.4) was not

[1]These data were analyzed by Clara DeLeon.

significantly different from that obtained by the whites (26.0). Under all other conditions of administration the mean scores of the black students was significantly different from that of the white students.

The finding of no performance differential between blacks and whites under one set of conditions in which KR was given is certainly an important one, considering the significance of the main effect found for race and the widespread finding of lower ability test performance levels for blacks (e.g., Loehlin, Lindzey & Spuhler, 1975). However, the result was found only in the one order condition and thus is difficult to interpret. Further analysis, however, revealed that the results might be attributed to motivational effects. Under KR conditions when the conventional test had been administered first, blacks omitted almost no items, while under other conditions they omitted more items than whites. The results of this study must be interpreted with caution, however, because of the small total sample size and the small number of black students.

Summary

The limited number of studies available on the effects of KR on test performance yield conflicting findings. Studies reported by Beeson (1973), Heald (1970), Sweet and Ringness (1971), and Betz (1975) suggest that ongoing KR may facilitate test performance, although the latter two studies found interactions between the effects of KR and racial/socioeconomic variables. Studies by Bierbaum (1965), Spencer and Barker (1969), and Strang and Rust (1973) indicated that examinees made more errors under KR conditions, while the study by Zontine et al. (1972) found no differences in performance as a function of KR.

However, the generalizability of these findings is limited. Most testing today of ability and intelligence is done using standardized objective tests, yet almost all of the evidence relevant to the use of KR on objective tests comes from studies using unstandardized classroom achievement tests. The lack of standardization in such tests and the variety of approaches to their construction may explain the conflicting research findings.

Studies using classroom achievement tests also can be criticized for their failure to control the medium of test administration and/or mode of test response (Sympson, 1975). In most studies reported (e.g., Heald, 1970; Pressey, 1950; Spencer & Barker, 1969), the test using KR has been adminstered using some type of punchboard device, while the test not using KR has been administered using a standard (e.g., IBM) answer sheet. It is possible that observed performance differences in such cases may be due partly to different amounts of time taken to respond to items presented under the two different formats, differing amounts of effort or interest on the part of the testees, or unfamiliarity with the testing equipment.

In addition to the lack of generalizability and the failure to control the medium for responding to the test, too little attention has been paid to how the effects of KR may be moderated by other characteristics of the examinees or of the tests being administered. For example, two studies (Sweet & Ringness, 1971; Betz, 1975) found that the effects of KR were moderated by race and/or social class variables. Sweet and Ringness hypothesize that upper and middle class individuals may be maximally motivated to do their best and thus may not need the positive motivating effects of KR.

Finally, no study to date has given attention to the fact that the *quality* of the KR given, that is, the extent to which it is predominantly positive or negative, may influence its effectiveness as a motivational factor. Conventional ability tests are constructed to be maximally appropriate in difficulty level to the ability levels of average individuals in a group. But in these tests, the quality of the KR varies directly with the ability level of the examinee. That is, high-ability examinees receive mostly positive (i.e., "correct") KR, average-ability examinees receive about half positive and half negative (i.e., "incorrect") KR, and low-ability examinees receive mostly negative KR. On a conventional test, therefore, high-ability examinees are likely to be encouraged, and thus perhaps motivated, when they are provided with KR; but low-ability examinees may be discouraged and frustrated, rather than motivated, when they are provided with KR.

On an adaptive ability test, on the other hand, the items are selected to be appropriate in difficulty to each individual's ability level rather than to the mean ability level of some group of examinees, as in conventional testing procedures. An adaptive test is constructed so that each examinee answers correctly about half of the items administered; therefore, all examinees, regardless of ability level, should receive about half positive and half negative KR. Consequently, the proportion of positive KR on an adaptive test is relatively constant across individuals of different ability levels and its effects may be different from those observed on a conventional test.

Purposes of the Present Study

The purpose of the present study was to examine the effects of immediate knowledge of results on a computer-administered test of verbal ability. An additional focus of the study was to determine whether or not the effects of KR differed for conventional and adaptive tests, or for testees of different ability levels.

While the major dependent variable of interest was level of performance on the ability test, the effects of KR on two other aspects of test-taking behavior, response latency and response consistency, also were studied. In addition, the effects of KR versus no-KR conditions on psychometric characteristics of the conventional test were examined. The study also was concerned with the duration of KR effects, in terms of whether receiving KR had any effects on performance on a test given subsequently under no-KR conditions.

METHOD

Design

Independent Variables

This study utilized a randomized block analysis of variance design with three independent variables. The blocking variable was subject group. Groups were high-ability and low-ability college students. Within each group each student was randomly assigned to one of four treatment combinations, resulting from the cross-classification of two conditions of knowledge of results (KR), and two different strategies of measuring ability. In one condition of the

KR factor, examinees were informed after each response whether their responses were correct or incorrect. If the response was incorrect, they were informed of the correct multiple-choice alternative. In the other condition, examinees did not receive KR.

One strategy of measuring ability was a 50-item conventional ability test. In this test item difficulties were concentrated at median ability level of the high-ability group. The other test was an adaptive ability test, in which the items were selected to be appropriate to each individual's ability level. The adaptive testing strategy used was the stradaptive test (Weiss, 1973).

Dependent Variables

The primary dependent variable of interest was performance level on the ability test. Two methods of scoring the conventional test and two methods of scoring the stradaptive test were used to obtain estimates of ability. Alternate scoring methods were used to determine whether the obtained pattern of results differed as a function of the methods of scoring the tests.

Response latency was also a dependent variable of interest. Response latency was measured as the elapsed time from the presentation of a test item until the testee responded to the item. Response latencies were analyzed to determine whether they were affected by the provision of KR.

A third dependent variable was response consistency. In addition to providing estimates of ability level, the stradaptive testing strategy yields measures of the *consistency* of an individual's responses to test items (Weiss, 1973, pp. 26-27; 1974, pp. 52-53). Response consistency in an ability test reflects the range of confidence which can be attributed to a given estimate of ability level. Indices of response consistency were used to determine whether examinees responded in a more consistent manner under KR than under no-KR conditions.

To study the effects of providing KR on the psychometric properties of the conventional test, its internal consistency reliabilities within KR and no-KR conditions were compared.

While the subjects studied by Pressey (1950) reported that they liked receiving KR, they also indicated feeling frustrated when taking tests on which they no longer received it. Since frustration or other reactions to changed conditions may influence test perform[illegible] the design of the present study permitted the investigation of whether [illegible]eceiving KR had any effects on performance on a test given immedia[illegible] afterwards under standard (i.e., No-KR) administrative conditions.

Test Construction

Item Pool

The item pool used to construct the conventional and stradaptive tests of verbal ability consisted of five-alternative multiple-choice vocabulary items. The items were normed on University of Minnesota students, most of whom were from the College of Liberal Arts (see McBride & Weiss, 1974). Normal ogive difficulty (b) and discrimination (a) parameter estimates were available for each item. The pool contained about 400 vocabulary items that had a values greater than or equal to .30. The difficulty levels of these items were distributed across the continuum of underlying ability, with most values falling between ±3 standard units.

Stradaptive Test

Item structure and branching. For construction of the stradaptive test, the items in the pool were grouped into nine levels, or strata, on the basis of their difficulties. (See Appendix Table A-1 for the difficulties and discriminations of all items in the stradaptive test.) Each stratum included items whose range of difficulty (i.e., the difference in difficulty values between the most and the least difficult items in the stratum) was .67. There was no overlap in item difficulties between adjacent strata. Items ranged in difficulty from $b=-3$ to $b=+3$.

Once items had been grouped into difficulty levels, they were selected for inclusion in the test on the basis of their discriminating power. For any one stratum, the most highly discriminating item was selected first, and each successive item chosen had a lower discrimination. In this way, 30 items were selected for each stratum for which there were sufficient items available in the pool. However, no item having a discrimination less than $a=.30$ (which corresponds approximately to a biserial item-total score correlation of .28) was considered acceptable; as a result the strata at the extreme levels of difficulty did not contain 30 items. The smallest number of items in a stratum was 17. A total of 243 items comprised the stradaptive item structure.

Entry into the stradaptive test was determined on the basis of the examinee's self-reported grade-point average. Appendix Table A-2 indicates the entry stratum corresponding to each of nine GPA intervals. Those examinees reporting high GPAs began the test with more difficult items than did those reporting lower GPAs. Examinees were branched through the stradaptive item structure according to the rule that following a correct response, the most discriminating item remaining in the next more difficult stratum was administered, and following an incorrect response, the most discriminating item in the next less difficult stratum was administered.

Testing was terminated when either a ceiling stratum had been identified or 75 items had been administered. Since the items used were five-alternative multiple-choice items, the ceiling stratum was defined as that stratum where the examinee answered 20% or fewer of the items correctly, based on a minimum of five items administered at that stratum. However, there were some examinees whose response patterns never permitted the identification of a ceiling stratum. This could happen only for very high-ability examinees capable of responding at better than chance level at even the most difficult stratum. If a ceiling stratum had not been identified after the administration of 75 items, testing was terminated.

Scoring. Both ability level scores and consistency scores were calculated for stradaptive test response protocols.

In the stradaptive test, examinees answer different numbers of items, and the items that they answer vary in difficulty according to the individual's ability level. Thus, simple number-correct scores are not appropriate as ability estimates. However, maximum likelihood scores (Birnbaum, 1968) are appropriate because they take into account the difficulty and discrimination of each item administered and because they do not depend on the number of items administered to an individual. Accordingly, maximum likelihood scores were calculated for the stradaptive test.

The likelihood equation for the 3-parameter logistic model given by Birnbaum (1968, p. 459) was solved for the maximum likelihood estimate of each examinee's ability. Difficulty and discrimination parameters used for each item administered are those given in Appendix Table A-1. The guessing parameter (c) was set at .20 since each item had five response alternatives. Input into the scoring program consisted of each examinee's vector of 1's and 0's, corresponding to correct and incorrect responses respectively, along with the corresponding item parameters.

Ten simpler methods of scoring the stradaptive test were proposed by Weiss (1973, pp. 20-26). However, results reported by Vale & Weiss (1975a,b) indicated that the average difficulty of all items answered correctly (Score 8) was the best of the ten originally proposed methods of scoring the stradaptive test. This score requires fewer assumptions than the maximum likelihood score and considerably less computational time. Consequently, it was used as a dependent variable in this study to determine whether its results were the same as those obtained from maximum likelihood scoring.

Weiss (1973) suggested that the consistency of a response pattern might be related to the confidence with which ability is measured by a given set of test items. Consistency of response for an individual is to some extent analogous to discrimination indices characterizing items. An item discrimination index reflects the extent to which people having high levels of the trait of interest respond correctly to an item more often than do people having lower levels of that trait. Similarly, individuals should respond correctly to easier items more often than they respond correctly to more difficult items. If individuals answer many easy items (i.e., items below their ability levels) incorrectly and many difficult items (i.e., items above their ability levels) correctly, they are responding inconsistently, and it may be inferred that something besides the trait of interest is influencing their responses. In general, consistent testees are those whose response records contain less variability in the difficulties of items they encounter and answer correctly. More consistent testees will also answer items drawn from a smaller number of strata.

Weiss (1974, pp. 52-53) suggested five different consistency scores for use with the stradaptive test. Research by Vale and Weiss (1975a,b) and analyses of the present data (Betz, 1976) indicated that there are two clusters in these consistency scores; consequently, one score was selected as representative of each cluster. Consistency Score 1 (Score 11 in Vale & Weiss, 1975a,b) is defined as the standard deviation of the difficulties of all items encountered by a testee. Consistency Score 2 (Score 15 in Vale & Weiss) is the number of strata between the basal and ceiling strata. This score corrects for inappropriate entry points, or entry strata which are below the basal stratum or above the ceiling stratum.

In summary, maximum likelihood ability level scores, an average difficulty ability score, and two consistency scores were selected for analysis of performance and test-taking behavior on the stradaptive test. The maximum likelihood score was comparable to that used for the conventional test and thus permitted direct inter-strategy comparisons. The remaining scores were unique to the stradaptive test and therefore were analyzed only within that testing strategy.

Peaked Conventional Test

The peaked conventional test consisted of 50 items with difficulty values concentrated around b=-.20 and discrimination values greater than or equal to a=.40. The characteristics of the 50 items constituting the test are summarized in Table 1. While the mean difficulty value was b=-.20, the easiest item had b=-.97, and the most difficult item had b=.56. The average item discrimination ($\bar{a}$=.89) was considerably higher than the minimally acceptable level (a=.40), but there was considerable variation among items. Appendix Table A-3 provides the normal ogive difficulty and discrimination parameter values characterizing each item in the test. Items were administered in the order shown in Appendix Table A-3.

Table 1
Summary of the Characteristics
of Items in the 50-Item Conventional Test

Item Difficulty (b)				Item Discrimination (a)			
Mean	S.D.	Minimum	Maximum	Mean	S.D.	Minimum	Maximum
-.20	.38	-.97	.56	.89	.34	.41	1.90

The conventional test was scored using simple number-correct scores and maximum likelihood scores based on Birnbaum's (1968) three-parameter logistic model.

Post-Test

To determine whether there were any carry-over effects on later test performance for students who had received KR on the initial test, a 44-item post-test was administered to all testees following the administration of the experimental (i.e., peaked conventional or stradaptive) test. This test was constructed by selecting items from a pool of 120 vocabulary items from the Cooperative School and College Ability Tests,[2] forms 2A, 2B, 3A, and 3B. The items, like those in the item pool used for the stradaptive and peaked conventional test, were five-alternative multiple-choice vocabulary items. They were normed in a population of high school students, and normal ogive difficulty and discrimination parameters were available for each item. The test was con-

[2]These items were made available for research use by Educational Testing Service.

structured to have a rectangular distribution of item difficulties; that is, item difficulties were spaced approximately evenly across the ability/difficulty continuum and thus included very easy to very difficult items.

Table 2 shows the mean and standard deviation of the normal ogive difficulty and discrimination values characterizing the 44 items in the test. While the mean difficulty of these items ($\bar{b}$=-.19) was almost identical to that of the items in the peaked conventional test ($\bar{b}$=-.20), the normative populations from which the item parameters were derived differed substantially (i.e., high school students for the post-test parameters and college students for the experimental test parameters). It was expected, therefore, that the post-test items would be easier for college students (the population of interest in the present study) than would be items from the peaked test having numerically comparable item difficulty values.

Table 2
Summary of the Characteristics
of the Items in the 44-Item Post-Test

Item Difficulty (*b*)				Item Discriminatic (*a*)			
Mean	S.D.	Minimum	Maximum	Mean	S.D.	Minimum	Maximum
-.19	1.37	-2.85	2.62	1.22	.40	.51	1.94

Appendix Table A-4 provides normal ogive difficulty and discrimination parameters for each of the 44 items in the post-test. Items were administered in the order indicated in Table A-4. Number-correct scores were determined for each testee.

Procedure

Subjects

Two groups of students participated in this study. The first group consisted of 239 students taking the introductory psychology course in the College of Liberal Arts (CLA) at the University of Minnesota. The second group consisted of 111 students from psychology courses in the University's General College (GC). Both received two points toward their final course grade for participation in the experiment. The CLA students were considered a *High-Ability* group, i.e., a group consisting of people who typically perform relatively well on ability and scholastic aptitude tests. General College has lower admission standards than does CLA. Thus the GC students comprised the *Low-Ability* group, based on their lower mean ability level on standard tests of ability and scholastic aptitude.

Test Administration

All students were tested at individual cathode-ray terminals (CRTs) connected to a Hewlett-Packard 9600E Real-Time computer system. Test items were presented at 960 characters per second on the CRT screen, and testees indicated their responses by typing in the number corresponding to the chosen alternative for each five-alternative multiple-choice item. Instructional screens explaining the operation of the CRTs were provided prior to testing (see DeWitt & Weiss, 1974, pp. 36-53), and a proctor was present in the testing room to provide assistance to any testee having difficulty with the equipment or instructions. Students were permitted as much time as necessary to complete the tests and were so informed before testing was begun.

Experimental treatment. Immediate knowledge of results was provided to one half of the examinees. After the examinee responded to the test item, a message appeared on the screen below the item just answered. A correct response to the item was followed by the message, "That's correct". An incorrect response was followed by the message, "That's not correct. The correct answer is x," where x was the number corresponding to the correct multiple-choice alternative. In both cases, the testees then were allowed to examine the item and were to press the "return" key when they were ready for the administration of the next item. In the groups that did not receive KR, a new item was presented immediately following the examinees' responses to the previous item.

Testing sequence. After examinees had completed the instructional screens and had answered several identification and demographic questions, test administration was begun. First, either the 50-item peaked conventional test or the stradaptive test was administered with or without KR. Second, testees were administered several items concerning their reactions to the testing situation and, in the KR group only, their reactions to the provision of immediate knowledge of results (analyses of these data are reported by Betz & Weiss, 1976). Following completion of the reaction items, all examinees were administered the 44-item post-test.

Data Analysis

Several types of data were available for all individuals participating in the study, while other data were available only for testees completing either the stradaptive or the conventional experimental test. Data available for all testees included: 1) maximum likelihood ability estimates (scores) for the experimental test; 2) post-test number-correct scores; and 3) response latency data for each item administered. Data available for subgroups of testees included: 1) number-correct scores for examinees completing the peaked conventional experimental test; and 2) the average difficulty score and two consistency scores for examinees completing the stradaptive experimental tests.

Analysis of Ability Estimates

Mean differences. Maximum likelihood ability estimates obtained from the conventional and stradaptive tests were analyzed using a three-way analysis of variance. The three factors--KR, testing strategy, and ability group--were completely crossed and each had two levels. Because cell frequencies in the three-way crossed classification were neither exactly equal nor proportional, it was necessary to use computational procedures in the analysis of variance to account for the lack of orthogonality among main effects and between main and interaction effects. Computations were based on the "classic experimental" approach described by Nie, Hull, Jenkins, Steinbrenner, and Bent (1975, pp. 405-408).

Within each of the eight groups resulting from the 2 x 2 x 2 design, the mean and standard deviation of scores were calculated. Since each experimental variable consisted of two levels, significant main effects indicated a significant difference between the two means involved. To determine which combinations of testing conditions resulted in significantly high or low test performance, comparisons of the subgroup means were made using Scheffé's (1959) method.

Two-way analyses of variance and post-hoc comparisons also were used for the analysis of ability estimates obtainable from only one of the two experimental tests. For the peaked conventional test the number-correct score was analyzed, and for the stradaptive test the average difficulty score was analyzed.

Internal consistency reliability. The internal consistency reliability of the peaked conventional test was calculated using Cronbach's (1951) alpha formula for the total group of examinees taking the test and separately within the KR and No-KR subgroups. The significance of the difference between the reliability coefficients under KR and No-KR conditions was calculated using the formula suggested by Glass and Stanley (1970, p. 311). The formula is based on Fisher's Z transformation of r and was applied to the alpha reliability values, r_1 and r_2, after conversion to Z_{r_1} and Z_{r_2}.

Other Response Characteristics

Response consistency. Two-way analyses of variance of each of the two consistency scores derived from the stradaptive test records were completed using ability group and KR as the independent variables. Mean scores within each treatment combination were calculated, and *a posteriori* contrasts were studied.

Response latencies. The response latency for each item administered to each individual was available for both the stradaptive and conventional tests. Response latencies were recorded from the time the display of a test item was begun until the examinee pressed the "return" key to record his answer to the item. Latencies, in seconds, were accurate to 1/10 second. The mean response latency over all items administered was calculated for each testee, thus yielding a latency "score" for each individual. Latency scores were analyzed using a three-way analysis of variance.

Carry-Over Effects

Post-test scores. Within the KR and No-KR treatment groups, the number-correct scores on the post-test were analyzed using a three-way analysis of variance, with KR, testing strategy and ability group as the independent variables. Means and standard deviations of scores were calculated within each treatment-subject group combination, and contrasts on the means were made.

Correlation of experimental and post-test scores. To determine whether KR affected the relative positions of individuals within a group, correlations between post-test number-correct scores and the experimental test maximum likelihood ability estimate were calculated. These correlations were calculated separately for groups completing conventional and stradaptive tests for both KR and No-KR conditions. To determine whether there were greater differences between experimental test scores and post-test scores for the KR and No-KR conditions, the differences among the four correlation coefficients were tested for statistical significance using the procedure suggested by Glass and Stanley (1970, pp. 311-313).

RESULTS

Analysis of Ability Estimates

Maximum Likelihood Scores

Table 3 shows the results of the three-way analysis of variance of the maximum likelihood ability estimates obtained from the conventional and stradaptive tests. Table 3 also indicates the numbers of examinees and the means and standard deviations of scores associated with each treatment combination and for combined treatments. As the table indicates, there were significant main effects for Ability Group and for KR, but there were no significant interaction effects. Only the interaction between Ability Group and test approached statistical significance ($p=.122$).

As shown in Table 3, the overall mean level of performance of the High-Ability group (-.26) was significantly ($p<.01$) higher than that of the Low-Ability group (-.87), supporting their *a priori* ability level designations. Table 3 also shows that the performance level of both groups was significantly ($p<.05$) higher under KR conditions than under No-KR conditions.

Figure 1 shows a plot of the means for the eight experimental groups. Contrasts on the means for the eight subgroups indicated that there were three subgroups of means which were not significantly different within subgroups. The dashed lines in Figure 1 differentiate those three subgroups.

As Figure 1 shows, in both subject groups performance on the conventional test was significantly better under KR conditions; the High-Ability-KR mean (-.06) was significantly greater than the High-Ability-No-KR mean (-.43), and the Low-Ability-KR mean (-.87) was significantly higher than the Low-Ability-No-KR mean (-1.20). On the stradaptive test, however, the level of performance of the High-Ability group under KR conditions (-.19) was significantly greater than that under No-KR conditions (-.39), but the differences for the Low-Ability group (-.69 and -.72) and for the combined groups (i.e., High-Ability and Low-Ability) were not statistically significant.

Figure 1 also shows that significant differences between ability level groups were not found under all testing conditions. Although the overall level of performance in the High-Ability group was significantly higher than that of the Low-Ability group (-.26 vs. -.87), the performance of the Low-Ability group on the stradaptive and conventional tests under KR conditions was not

Table 3
Means and Standard Deviations of Maximum Likelihood Ability Estimates for Conventional and Stradaptive Tests in High- and Low-Ability Groups With and Without KR, and Three-Way ANOVA Results

	Experimental Condition						Combined Conditions		
	KR			No-KR					
Test and Group	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.
Conventional Test									
High-Ability	60	-.06	1.04	57	- .43	1.22	117	- .24	1.14
Low-Ability	28	-.87	.84	28	-1.20	1.40	56	-1.03	1.16
Stradaptive Test									
High-Ability	60	-.19	1.21	62	- .39	.91	122	- .29	1.07
Low-Ability	27	-.69	.79	27	- .72	.89	55	- .71	.83
Combined Groups									
Conventional Test	88	-.31	1.05	85	- .68	1.33	173	- .49	1.20
Stradaptive Test	87	-.35	1.12	89	- .49	.91	176	- .42	1.02
High-Ability	120	-.12	1.13	119	- .41	1.07	239	- .26	1.10
Low-Ability	55	-.78	.82	55	- .97	1.20	110	- .87	1.02
Total Group	176	-.33	1.09	174	- .58	1.14	349	- .46	1.11

Three-Way Analysis of Variance

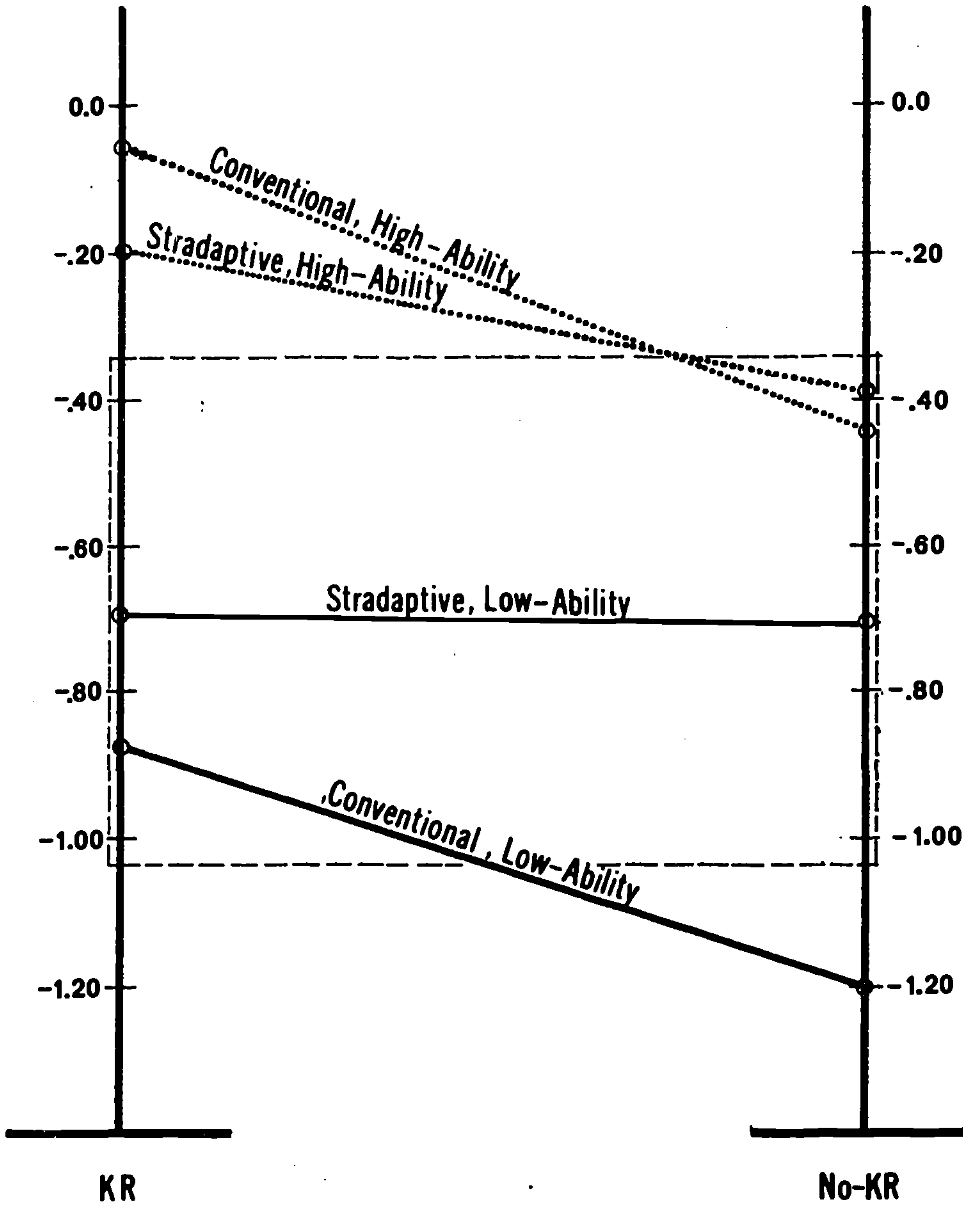
Source of Variation	Sum of Squares	DF	Mean Square	*F*	*p*[1]
Main Effects	33.84	3	11.28	9.79	.001
Ability Group	27.67	1	27.66	23.99	.001
Test	.42	1	.42	.36	.999
KR	5.63	1	5.63	4.88	.026
Two-Way Interactions	3.90	3	1.30	1.13	.340
Ability Group x Test	2.71	1	2.71	2.35	.122
Ability Group x KR	.17	1	.17	.15	.999
Test x KR	1.02	1	1.02	.89	.999
Three-Way Interaction					
Ability Group x Test x KR	.07	1	.07	.06	.999
Residual	393.15	341	1.15		
Total	430.96	348	1.24		

[1]Estimated probability of error in rejecting null hypotheses.

significantly lower than that of the High-Ability group under No-KR conditions on both tests. It may be noted further that the performance of the High-Ability group was highest under KR conditions, while the performance level of the Low-Ability group was high under KR conditions *and* on the stradaptive test in general.

From these results it appears that the performance of the High-Ability group was enhanced when KR was given regardless of testing strategy. On the other hand, performance of the Low-Ability group was improved under either KR

Figure 1
Mean Maximum Likelihood Ability Estimates
as a Function of Testing Strategy, KR, and Ability Group



conditions or by administration of an adaptive test. When no KR was provided on the conventional test--the conditions typical of most standard testing procedures--the performance level of Low-Ability individuals was significantly lower than that of the High-Ability individuals. But this performance differential between the two groups did not appear, under No-KR conditions, using an adaptive test.

Other Ability Scores

Conventional test. The means for the number-correct scores within each KR-subject group combination and the results of the two-way analysis of variance of the means are shown in Table 4. These results show significant main effects for both Ability Group and KR, but no interaction effects. Again, the performance of the High-Ability group was significantly better than that of the Low-Ability group, and the overall mean score under KR conditions was higher than that under No-KR conditions. However, contrasts on the subgroup

Table 4
Means and Standard Deviations of Number Correct Scores on the 50-Item Conventional Test for Two Ability Level Groups With and Without KR, and Two-Way ANOVA Results

	Experimental Condition						Combined Conditions		
	KR			No-KR					
Group	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
High-Ability	60	30.47	9.20	57	27.10	10.31	117	28.83	9.90
Low-Ability	28	22.54	8.28	28	20.71	9.39	56	21.62	8.90
Total	88	27.94	9.61	85	25.00	10.41	173	26.50	10.09

Two-Way Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	F	p[1]
Main Effects	2319.82	2	1159.91	12.92	.001
Ability Group	1945.29	1	1945.29	21.67	.001
KR	354.28	1	354.28	3.95	.046
Ability Group x KR	22.45	1	22.45	.25	.999
Error	15170.98	169	89.77		
Total	17513.25	172	101.82		

[1]Estimated probability of error in rejecting null hypotheses.

means indicated that while the High-Ability-KR mean (30.47) was significantly greater than the Low-Ability-KR (22.54) mean or the Low-Ability-No-KR (20.71) mean, the High-Ability-No-KR mean (27.10) was not significantly different from the Low-Ability-KR mean (22.54).

Stradaptive test. Mean average difficulty scores as a function of KR and subject group and the results of the two-way analysis of variance are in Table 5. Only the Ability Group effect was significant in these data. As expected, the scores of the High-Ability group were higher than those of the Low-Ability group. Contrasts among the subgroup means indicated that there were no further significant mean differences. This finding of no KR or interaction effects is in agreement with the results shown for the stradaptive maximum likelihood scores within the Low-Ability group; but it does not agree with the finding of significant KR effects for the High-Ability group using maximum likelihood scores.

Table 5
Means and Standard Deviations of Average-Difficulty-of-Items-Answered-Correctly Scores on the Stradaptive Test for Two Ability Level Groups With and Without KR, and Two-Way ANOVA Results

	Experimental Condition						Combined Conditions		
	KR			No-KR					
Group	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.
High-Ability	54	-.21	1.00	59	-.38	.89	113	-.30	.95
Low-Ability	28	-.67	.82	26	-.61	.92	54	-.64	.86
Total	82	-.37	.96	85	-.45	.90	167	-.41	.94

Two-Way Analysis of Variance

Source of Variation	Sum of Squares	DF	Square	*F*	*p*[1]
Main Effects	5.28	2	2.67	3.05	.05
Ability Group	4.20	1	4.20	4.86	.03
KR	1.38	1	1.38	1.59	.21
Ability Group x KR	.81	1	.81	.94	.99
Error	121.94	141	.86		
Total	128.02	144			

[1]Estimated probability of error in rejecting null hypotheses.

Internal Consistency Reliability

Coefficient alpha for the 50-item conventional test was .90 when calculated for the total group of examinees taking the test. The reliability of the test under KR conditions was .89, while that under No-KR conditions was .91. The difference between the reliability coefficients for KR and No-KR conditions was not statistically significant.

Other Response Characteristics

Response Consistency

Tables 6 and 7 show means for the two stradaptive consistency scores, by Ability Group and KR conditions, and the results obtained from the two-way analyses of variance for each score. There were no significant main or interaction effects for either of the scores, nor were there significant differences among any of the cell means. Thus, response consistency was not significantly influenced by either KR conditions or by ability level of the testees.

Table 6
Means and Standard Deviations of Stradaptive Consistency Score 1 as a Function of KR Condition and Ability Level Group, and Results of the Two-Way ANOVA

	Experimental Condition						Conditions Combined		
	KR			No-KR					
Group	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.
High-Ability	54	.78	.18	59	.74	.17	113	.76	.18
Low-Ability	28	.80	.22	26	.85	.27	54	.82	.26
Total	82	.78	.19	85	.77	.19	167	.78	.19

Two-Way Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	*F*	*p*[1]
Main Effects	.07	2	.04	1.14	.32
Ability Group	.04	1	.04	1.30	.25
KR	.02	1	.02	.83	.99
Ability Group x KR	.06	1	.06	1.99	.16
Residual	4.31	141	.03		
Total	4.41	144	.03		

[1]Estimated probability of error in rejecting null hypotheses.

Table 7
Means and Standard Deviations of Stradaptive Consistency Score 2 as a Function of KR Condition and Ability Level Group, and Results of the Two-Way ANOVA

	Experimental Condition						Conditions Combined		
	KR			No-KR					
Group	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.
High-Ability	54	1.59	1.25	59	1.83	1.15	113	1.72	1.20
Low-Ability	28	1.64	1.22	26	1.23	1.11	54	1.44	1.19
Total	82	1.61	1.23	85	1.65	1.16	167	1.63	1.19

Two-Way Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	*F*	*p*
Main Effects	1.36	2	.68	.58	.99
Ability Group	1.35	1	1.35	1.14	.29
KR	.05	1	.04	.03	.99
Ability Group x KR	2.11	1	2.11	1.79	.18
Residual	166.29	141	1.18		
Total	169.77	144	1.18		

Response Latency

Means and standard deviations of response latency scores as a function of KR, Test, and Ability Group, and the results of the three-way analysis of variance of mean latency scores, are shown in Table 8. Table 8 indicates that the only significant main effect was for Ability Group. High-Ability examinees took significantly less time to respond to test items than did Low-Ability examinees; the mean response time for the former group was 14.9 seconds while that of the latter group was 16.7 seconds. Response latency did not differ significantly as a function of Test or KR, and there were no significant interaction effects.

Table 8
Means and Standard Deviations for Average Intra-Individual Response Latency in Seconds, and Three-Way ANOVA Results

	Experimental Condition				Combined Conditions	
	KR		No-KR			
Test and Group	Mean	S.D.	Mean	S.D.	Mean	S.D.
Conventional Test						
High-Ability	14.4	4.5	14.7	5.0	14.6	4.8
Low-Ability	15.2	4.4	17.4	6.9	16.3	5.8
Stradaptive Test						
High-Ability	15.2	5.3	15.2	5.1	15.2	5.2
Low-Ability	18.0	9.1	16.1	5.5	17.1	7.6
Combined Groups						
Conventional Test	14.7	4.5	15.6	5.8	15.1	5.2
Stradaptive Test	16.1	6.8	15.4	5.2	15.8	6.1
High-Ability	14.8	4.9	14.9	5.1	14.9	5.0
Low-Ability	16.6	7.2	16.7	6.2	16.7	6.7
Total	15.4	5.8	15.5	5.5	15.4	5.6

Three-Way Analysis of Variance

Source of Variation	DF	Mean Square	F	p[1]
KR	1	1.18	.04	.999
Test	1	40.72	1.30	.254
Ability Group	1	246.82	7.87	.006
KR x Test	1	47.56	1.52	.217
KR x Ability Group	1	.03	.001	.999
Test x Ability Group	1	.35	.01	.999
KR x Test x Ability Group	1	69.16	2.20	.135
Residual	342	31.37		

[1]Estimated probability of error in rejecting null hypotheses.

Carry-Over Effects

Post-Test Scores

Means and standard deviations of post-test number-correct scores as a function of Ability Group and KR conditions on the experimental test are shown in Table 9; the table also shows the results of the three-way analysis of variance of mean post-test scores. As shown in Table 9, there was a significant main effect for Ability Group; the mean number correct in the High-Ability group was 35.6, or about 81% correct, while that in the Low-Ability group was 32.5, or about 74% correct. There were no other significant main or interaction effects, indicating that performance on the post-test was not affected by differences in the conditions under which the experimental test was administered. Thus, while testing conditions did influence test performance while they were in effect, there were no discernible carry-over effects on test performance on a conventional test administered immediately after the experimental test.

Table 9
Means and Standard Deviations of Number-Correct Scores on the Post-Test as a Function of Experimental Conditions, and Three-Way ANOVA Results

	Experimental Condition				Combined Conditions	
	KR		No-KR			
Test and Group	Mean	S.D.	Mean	S.D.	Mean	S.D.
Conventional Test						
High-Ability	36.23	5.05	35.47	4.52	35.86	4.79
Low-Ability	32.93	6.15	32.21	6.08	32.57	6.07
Stradaptive Test						
High-Ability	35.38	5.19	35.31	5.53	35.34	5.34
Low-Ability	32.86	6.95	32.07	6.37	32.47	6.62
Combined Groups						
Conventional Test	35.18	5.60	34.40	5.28	34.80	5.45
Stradaptive Test	34.58	5.89	34.33	5.95	34.45	5.90
High-Ability	35.81	5.11	35.39	5.05	35.60	5.07
Low-Ability	32.89	6.50	32.14	6.17	32.52	6.32
Total	34.88	5.74	34.36	5.62	34.62	5.68

Three-Way Analysis of Variance

Source of Variation	DF	Mean Square	F	p[1]
Ability Group	1	720.18	23.51	.001
Test	1	12.61	.41	.999
KR	1	23.72	.77	.999
Ability Group x Test	1	3.12	.10	.999
Ability Group x KR	1	2.07	.07	.999
Test x KR	1	4.32	.14	.999
Ability Group x Test x KR	1	2.68	.09	.999
Residual	342	30.64		

[1]Estimated probability of error in rejecting null hypotheses.

Correlation of Experimental and Post-Test Scores

For the conventional test group, the correlation between experimental-test maximum likelihood scores and post-test scores was higher, although not significantly, under No-KR conditions (r=.76) than under KR conditions (r=.69). On the stradaptive test, the correlation was again higher under No-KR conditions (r=.79) than under KR conditions (r=.76), but this difference also was not statistically significant. Thus, providing KR on a verbal ability test does not result in test scores which correlate substantially differently with scores on another test administered without KR than do scores obtained from individuals taking the same test under typical, i.e., No-KR, conditions of test administration.

SUMMARY AND CONCLUSIONS

Effects of KR on Test Performance

The results of the present study indicate that knowledge of results led to significant increases in test scores for the total group of examinees; that is, mean test scores were significantly higher under KR conditions than under No-KR conditions. However, the magnitude of the effects of KR on performance differed according to whether the test administered was a conventional or a stradaptive test.

The improvement under KR conditions was substantially greater for the conventional test than it was for the stradaptive test. Both the maximum likelihood and the number-correct scores on the conventional test were significantly higher under KR conditions than under No-KR conditions; this effect was significant for the total group of examinees and also within both the High-Ability and Low-Ability subgroups. While the KR score means were higher than the No-KR means for the stradaptive test scores, these differences were not significant for either the total group of examinees or for the Low-Ability group. Only the stradaptive test maximum likelihood scores in the High-Ability group were significantly higher under KR conditions.

Thus, providing KR on a conventional test of ability led to significant increases in mean test scores for both high- and low-ability testees. Providing KR on an adaptive test of ability led to increases in test scores for both ability-level groups, but the score increase was statistically significant only within the high-ability group.

These results indicate that KR alone can enhance ability test performance regardless of the ability level of the examinee, but only under conventional testing procedures. In contrast to the hypothesis of Sweet and Ringness (1971), the present study indicated that high-ability examinees achieved significantly higher scores under KR conditions even though they may be generally highly motivated to do well. Low-ability examinees were found to achieve significantly higher scores on the conventional test under KR conditions even though they generally received lower proportions of positive KR than did the high-ability examinees.

Group Differences and Testing Conditions

On both the conventional test and the stradaptive test, the High-Ability group obtained significantly higher mean scores than did the Low-Ability group. Thus, over all testing conditions combined, the performance differential between the two groups corresponded to that expected on the basis of their previous levels of performance on ability tests. However, and more importantly, results indicated that there were differences between the two ability-level groups in the effects of testing conditions on performance, and that under some conditions of test administration, the performance levels of the two groups were not significantly different.

The High-Ability group performed consistently and significantly better under KR conditions than under No-KR conditions on both the conventional and stradaptive tests, and there were no significant differences in this group between mean scores on the conventional and stradaptive tests. In contrast, while the Low-Ability students performed better on the conventional test under KR conditions, their performance on the stradaptive test did not differ as a function of KR conditions. Moreover, the performance of this group on the stradaptive test was consistently better than their performance on the conventional test even when the latter test had been administered under KR conditions. The score means for the Low-Ability group on the stradaptive test under both KR and No-KR conditions, and on the conventional test under KR conditions, were not significantly different from each other; but all three means were significantly higher than the group's mean on the conventional test under No-KR conditions. Further, the former three means in the Low-Ability group did not differ significantly from the means of the High-Ability group on either the conventional or stradaptive tests administered under No-KR conditions.

Thus it appears that the performance of low-ability examinees was enhanced *either* by providing these students with immediate knowledge of results *or* by administering to them an adaptive test of ability. These results imply that for low-ability students, adaptive testing might provide the same incentive effects as does the provision of KR.

Motivating Effects of Adaptive Testing

The incentive effects of an adaptive test for low-ability individuals may be because they perceive themselves as doing relatively well on an adaptive test in comparison to their usual performance on ability tests. Most group-administered ability and aptitude tests are constructed to be appropriate for individuals of average ability in the group for which the test is intended. Low-ability examinees probably perceive such tests as beyond their capabilities and may become discouraged early in the test. When these examinees have taken several tests that are too difficult for them, they may approach later testing situations with an expectation of further discouragement and failure.

However, on an adaptive test the items administered to low-ability examinees will be easier than those administered to average- or high-ability examinees. The stradaptive test is designed so that testees of all ability levels should be able to answer about half of the items administered to them correctly; and indeed, results indicated that the average examinee in the Low-Ability group obtained 46.5% correct, compared to about 40% correct on the conventional test with no KR. Consequently, low-ability examinees taking the stradaptive test probably perceived that, in relationship to their expectations, they were performing well on the test. It is possible that this situation served as an incentive for these individuals to try harder on the stradaptive test.

The absence of a motivating effect for the adaptive test in the High-Ability group may be explained by the same reasoning that explains a motivating effect in the Low-Ability group. The conventional test used in the present study was constructed to be maximally appropriate to individuals of about average ability in the normative population of high-ability students. On the basis of the mean difficulty level of the conventional test items, the average high-ability examinee was expected to answer about 54% of the items correctly. Results indicated that the average high-ability examinee obtained about 58% correct on the conventional test. Similarly, on the stradaptive test most examinees, regardless of ability level, should answer about 50% of the items administered to them correctly. In fact, the average High-Ability examinee obtained 50% correct on the stradaptive test.

Most high-ability students probably were accustomed to taking tests designed to be appropriately difficult for average individuals in their group. Therefore, the stradaptive test likely was perceived as an experience correspondent with their usual expectations of their level of test performance, and thus did not in itself have motivating effects. Undoubtedly there were some very high-ability students in the High-Ability group who perceived themselves as performing less well on the stradaptive test in comparison to their typical levels of ability test performance. But the possibly adverse effects for these students probably were balanced by the effects of the stradaptive test for some relatively low-ability students in the High-Ability group who, like most low-ability examinees, were pleasantly surprised by their levels of performance on the test.

On the conventional test the percentage correct for High-Ability testees (58%) was substantially greater than that for Low-Ability students (43%) and their mean ability scores differed substantially. On the other hand, on the stradaptive test the percentages of correct responses obtained by the two groups were more similar (50% vs. 46.5%), and their mean ability scores were also closer together.

The results concerning the effects of testing strategy and KR conditions on the performance of examinees of different ability levels are particularly important because of their implications for the measurement of "maximum performance" levels. Standard testing conditions (i.e., conventional objective tests administered without provision of KR) did not elicit maximum levels of performance from either group of examinees studied. Modifications of testing conditions, specifically the provision of KR for high-ability examinees and either the provision of KR or the administration of an adaptive test for low-ability examinees, were found to lead to significantly higher levels of performance. Perhaps more important, modifications of testing conditions usually assumed to elicit maximum levels of performance were found to reduce to insignificant levels, in some cases, score differences between two groups of supposedly different ability levels.

Other Effects of KR

The internal consistency reliability of the 50-item conventional test was not found to differ significantly as a function of KR conditions; reliability under KR conditions was .89, while that under No-KR conditions was .91. Thus, the data of the present study suggest that KR neither adds nor subtracts reliable variance in a set of ability scores. However, given the lack of other data relevant to this question, further study will be needed to delineate exactly what effects, if any, KR has on the precision of measurement.

Mean response consistency scores on the stradaptive test were not found to differ as a function of KR conditions. Thus, while KR increased mean test scores, it did not appear to increase the consistency of examinees' response patterns. Similarly, response latency was not related to KR conditions. Thus, KR did not affect the speed with which test responses were made.

The finding that there were no significant differences in mean response latency between the conventional and stradaptive tests differs from the findings reported by Waters (1975). Waters found that mean response times on the stradaptive tests were significantly longer than were response times on a conventional test; his interpretation of this finding was that examinees had to "think longer" about the answers to stradaptive test items because the items were selected to be at the limit of the examinee's ability level. In the present study, however, this appeared not to be the case. The differing result may have been due to characteristics of the test items or the testees. Further research on response latency differences between conventional and adaptive tests is indicated.

KR also did not systematically affect performance on a conventional test administered following the experimental test. Analysis of the post-test number-correct scores indicated that mean post-test scores did not differ as a function of previous testing conditions--either KR or test administered (i.e., the conventional or the stradaptive). From these data it may be concluded that the facilitating effects of KR on the experimental test did not transfer to performance on a test given subsequently under no-KR conditions. Further, if there were adverse reactions such as frustration to no longer receiving KR, these reactions did not affect performance to such a degree that examinees who had received KR obtained lower post-test scores than did those who had not received KR.

It is also interesting to note that the group effect was highly significant on the post-test. High-ability examinees obtained significantly higher scores than did low-ability examinees. However, results from the experimental tests had shown that under some conditions of administration, the scores of high- and low-ability examinees were not significantly different. The conditions in effect on the post-test were similar to those of most testing situations; knowledge of results was not provided and the test administered was a conventional rather than an adaptive test. Under these conditions, the group differences found corresponded to those expected on the basis of the previous levels of performance of the two groups.

KR also did not significantly affect the correlations of scores between the experimental test and a conventional post-test. Although these correlations were lower for the group of examinees who had recieved KR on the experimental test than for those who had not received KR, the differences between the correlations in the KR and no-KR groups were not statistically significant. Thus, although KR affected mean levels of test performance, effects for the total group were relatively constant across individuals. Further research utilizing repeated measures designs should be directed at investigating whether KR can result in significant individual differences in ability test performance.

Conclusions

The results of the present study demonstrated that providing examinees with immediate knowledge of results can lead to significant increases in ability test scores. Thus, it appears that knowledge of results can increase the extent to which ability tests measure the "maximum performance" capabilities of individuals. However, further research is needed to determine whether providing examinees with KR increases the validity of test scores.

In a group of low-ability examinees, test scores were higher on the stradaptive test than they were on the conventional test. This suggests that adaptive testing may have motivational effects similar to those of immediate knowledge of results, particularly for examinees for whom conventional tests are too difficult.

Testing conditions had somewhat different effects on the performance levels of high- and low-ability examinees, and there were some conditions under which the expected group differences in test scores were not found. This result suggests that testing conditions may affect not only the conclusions made about individuals on the basis of test scores, but the conclusions made about group differences in ability level. Therefore, in studying differences in psychological variables, more attention should be paid to the possible impact on the obtained results of the conditions under which measurements are made.

While knowledge of results can be provided on paper-and-pencil tests, its provision is, at best, inefficient and unwieldly. Consequently, it is not likely that providing KR will become standard in the administration of such tests. Further, most adaptive tests must be administered by computer. Studies of the few adaptive tests which can be administered by paper-and-pencil methods (e.g., the flexilevel test; Lord, 1971) have shown that significant numbers of examinees fail to follow the branching directions properly and thus invalidate their test protocols (e.g., Olivier, 1974).

It is evident that the administration of ability tests by computer provides psychological measurement with capabilities which have been either difficult or impossible to implement using paper-and-pencil testing methods. The facilitative effects of both immediate knowledge of results and adaptive testing on ability test performance found in the present study support the use of computer-assisted testing procedures to provide measurements consistent with a maximum performance conceptualization of human abilities.

REFERENCES

Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. Journal of General Psychology, 1956, 54, 279-299.

Annett, J. The role of knowledge of results in learning: A survey (Tech. Rep. 342-3). Orlando Fla.: U.S. Naval Training Device Center, 1961.

Annett, J. Feedback and human behavior. Baltimore: Penguin Books, 1969.

Bayroff, A. G. Feasibility of a programmed testing machine (Research Study 64-3). Washington, D.C.: U.S. Army Personnel Research Office, 1964.

Beeson, R. O. Immediate knowledge of results and test performance. Journal of Educational Research, 1973, 66, 224-226.

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. (AD A018675).

Betz, N. E. & Weiss, E. J. Psychological effects of immediate knowledge of results and adaptive testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976.

Bierbaum, W. B. Immediate knowledge of performance on multiple-choice tests. Journal of Programmed Instruction, 1965, 3, 19-23.

Bilodeau, E. A. & Bilodeau, I. McD. Motor-skills learning. In P. R. Farnsworth, O. McNemar & Q. McNemar (Eds.), Annual Review of Psychology (Vol. 12). Palo Alto, Calif.: Annual Reviews, 1961.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968, 397-479.

Brown, F. J. Knowledge of results as an incentive in schoolroom practice. Journal of Educational Psychology, 1932, 23, 532-552.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

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. (AD 773961).

DuBois, P. H. A history of psychological testing. Boston: Allyn & Bacon, 1970.

Ferguson, R. L. & Hsu, T. The application of item generators for individualizing mathematics testing and instruction (Report 1971/14). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1971.

Geis, G. L. & Chapman, R. Knowledge of results and other possible reinforcers in self-instructional systems. Educational Technology, 1971, April.

Glass, G. V. & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Heald, H. M. The effects of immediate knowledge of results and correlation of errors and test anxiety upon test performance. Doctoral dissertation University of Nebraska, 1970. (University Microfilms No. 70-17, 724).

Kulhavy, R. W. & Anderson, R. C. Delay-retention effect with multiple-choice tests. Journal of Educational Psychology, 1972, 63, 505-512.

Loehlin, J. C., Lindzey, G. & Spuhler, J. N. Race differences in intelligence. San Francisco, Calif.: W. H. Freeman, 1975.

Lord, F. M. The self-scoring flexilevel test. Journal of Educational Measurement, 1971, 8, 147-151.

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. (AD 781894).

McMahon, M. P. Effects of knowledge of ability test results on academic performance and test anxiety. Journal of Counseling Psychology, 1973, 20, 247-249.

Montor, K. Effect of using a self-scoring answer sheet on knowledge retention. Journal of Educational Research, 1970, 63, 435-437.

Newman, M. I., Williams, R. G., & Hiller, J. H. Delay of information feedback in an applied setting: Effects on initially learned and unlearned items. Journal of Experimental Education, 1974, 42, 55-59.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences. New York: McGraw Hill, 1975.

Olivier, P. An evaluation of the self-scored flexilevel tailored testing model. Unpublished doctoral dissertation, Florida State University, 1974.

Plowman, L., & Stroud, J. B. Effect of informing pupils of the correctness of their responses to objective test questions. Journal of Educational Research, 1942, 36, 16-20.

Pressey, S. L. A simple apparatus which gives tests and scores--and teaches. School and Society, 1926, 23, 373-376.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. Journal of Psychology, 1950, 29, 417-447.

Ross, C. The influence upon achievement of a knowledge or progress. Journal of Educational Psychology, 1933, 24, 609-619.

Samuda, R. J. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead & Co., 1975.

Sattler, J. M. & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 68, 347-360.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Spencer, R. E. & Barker, B. An applied test of the effectiveness of an experimental feedback answer sheet (Research Report No. 293). Urbana, Ill.: University of Illinois, Measurement and Research Division, 1969.

Strang, H. R. & Rust, J. O. The effects of immediate knowledge of results and task definition on multiple-choice answering. The Journal of Experimental Education, 1973, 42, 77-80.

Sympson, J. B. Problem: Evaluating the results of computerized adaptive testing. 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. (AD A018675).

Sweet, R. C. & Ringness, T. A. Variations in the intelligence test performance of referred boys of differing racial and socioeconomic backgrounds as a function of feedback or monetary reinforcement. Journal of School Psychology, 1971, 9, 399-408.

Terman, L. M. The measurement of intelligence. New York: Houghton Mifflin, 1916.

Terman, L. M. & Merrill, M. A. Stanford-Binet Intelligence Scale: Manual for the third revision Form L-M. Boston: Houghton Mifflin, 1960.

Vale, C. D. & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975 (a). (AD A018758).

Vale, C. D. & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975 (b). (AD A020961).

Waters, B. K. An empirical investigation of Weiss' stradaptive testing model. Proceedings of the first conference on computerized adaptive testing (Professional Series 75-6). Washington, D.C.: U.S. Civil Service Commission, Bureau of Policies and Standards, Personnel Research and Development Center, 1975. (Superintendent of Documents Stock No. 006-00940-9).

Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Weiss, D. J. Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1974. (AD A004270).

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (AD 768376).

Weiss, D. J. & Betz, N. E. Ability measurement: Conventional or adaptive? (Research Report 73-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (AD 757788).

Zontine, P. L., Richards, H. C. & Strang, H. R. Effect of contingent reinforcement on Peabody Picture Vocabulary Test performance. Psychological Reports, 1972, 31, 615-622.

Table A-1
Normal Ogive Difficulty (b) and
Discrimination (a) Parameters for Items
in the Stradaptive Test

	Stratum 1		Stratum 2		Stratum 3		Stratum 4		Stratum 5		Stratum 6		Stratum 7		Stratum 8		Stratum 9	
	b	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*
	-2.50	3.00	-2.13	3.00	-1.62	3.00	-.84	3.00	.11	1.90	.51	2.23	1.24	1.88	1.83	1.53	2.70	.86
	-2.50	3.00	-2.24	3.00	-1.26	3.00	-.70	3.00	.23	1.82	.75	1.56	1.28	1.81	1.90	1.06	2.48	.76
	-2.76	2.13	-1.87	3.00	-1.42	3.00	-.90	2.00	.16	1.64	.56	1.42	1.42	1.22	2.05	.82	2.45	.69
	-2.64	1.61	-2.24	3.00	-1.41	3.00	-.68	1.83	.30	1.47	.62	1.40	1.19	1.19	1.94	.78	2.57	.65
	-2.35	1.03	-2.20	3.00	-1.37	3.00	-.50	1.61	.18	1.42	.75	1.19	1.41	1.03	1.94	.74	2.58	.54
	-2.47	1.03	-1.75	3.00	-1.37	2.72	-.95	1.59	.13	1.41	.60	1.19	1.43	.89	1.90	.71	2.38	.54
	-2.40	.98	-1.87	2.89	-1.58	2.42	-.76	1.54	-.13	1.41	.60	1.18	1.60	.88	1.74	.71	2.91	.51
	-3.05	.91	-2.18	1.93	-1.16	2.31	-.63	1.50	-.33	1.30	.50	1.15	1.13	.78	2.29	.70	2.41	.48
	-2.66	.77	-1.83	1.93	-1.66	2.25	-.84	1.44	-.14	1.17	.36	1.10	1.07	.77	1.74	.66	2.61	.44
	-2.81	.76	-1.68	1.80	-1.18	1.99	-.36	1.29	-.22	.96	.59	1.02	1.49	.72	1.73	.65	2.83	.44
	-3.04	.71	-2.25	1.71	-1.21	1.83	-.64	1.18	-.16	.92	.56	1.00	1.32	.70	2.30	.58	2.57	.43
	-2.71	.67	-2.15	1.67	-1.63	1.73	-.93	1.16	-.26	.92	.72	.96	1.09	.65	2.20	.56	2.64	.40
	-2.59	.60	-2.19	1.56	-1.16	1.61	-.60	1.16	.07	.91	.66	.91	1.09	.65	1.85	.55	2.47	.40
	-2.93	.52	-1.90	1.46	-1.15	1.60	-.66	1.08	.29	.89	.76	.86	1.33	.64	1.89	.54	2.35	.38
	-2.38	.54	-1.70	1.24	-1.14	1.31	-.51	.97	-.22	.86	.34	.86	1.53	.60	2.09	.50	3.10	.36
	-2.77	.48	-1.90	1.22	-1.25	1.19	-.34	.96	.20	.85	.49	.85	1.38	.57	1.86	.49	2.90	.35
	-2.73	.46	-2.10	1.12	-1.03	1.17	-.36	.94	-.03	.84	.38	.74	1.36	.54	1.80	.48	2.84	.30
	-3.06	.37	-1.83	1.00	-1.36	1.11	-.36	.92	-.17	.81	.91	.72	1.57	.48	1.68	.48		
	-2.57	.36	-2.09	.97	-1.09	1.05	-.35	.86	.30	.78	.62	.73	1.22	.46	2.28	.47		
	-2.72	.34	-1.85	.97	-1.65	1.02	-.42	.83	-.30	.78	.56	.63	1.67	.38	2.24	.47		
	-2.37	.31	-1.82	.95	-1.24	.95	-.62	.79	-.24	.74	.42	.62	1.24	.38	2.22	.45		
			-1.82	.91	-1.04	.95	-.70	.77	.22	.74	.84	.58	1.65	.36	1.90	.43		
			-1.71	.87	-1.16	.89	-.46	.74	.32	.68	.86	.55	1.42	.33	2.19	.41		
			-1.86	.82	-1.10	.82	-.60	.71	.06	.64	.58	.54	1.52	.33	1.98	.40		
			-1.85	.79	-1.51	.77	-.70	.71	.01	.64	.68	.53	1.19	.32	1.72	.40		
			-2.18	.77	-1.49	.68	-.44	.71	-.02	.63	.57	.53	1.47	.32	2.28	.38		
			-1.96	.71	-1.03	.62	-.92	.65	.10	.52	.45	.52	1.43	.31	2.16	.32		
			-1.97	.67	-1.32	.56	-.73	.63	-.12	.43	.72	.45			2.15	.31		
			-2.26	.53	-1.08	.55	-.60	.58	-.13	.43	.35	.43						
			-2.31	.52	-1.33	.53	-.83	.50	-.23	.41	.48	.42						
	-2.67	.98	-1.99	1.57	-1.30	1.59	-.63	1.19	0	.96	.59	.90	1.36	.71	1.99	.59	2.63	.50
	.22	.80	.20	.88	.20	.87	.19	.63	.20	.41	.15	.41	.17	.42	.20	.25	.22	.16
ms	21		30		30		30		30		30		27		28		17	

Table A-2
Entry Strata for the Stradaptive Test
as a Function of Reported Grade-Point Average

Grade-Point Average	Entry Stratum	Mean Difficulty Level of Entry Stratum
3.76 to 4.00	9	2.63
3.51 to 3.75	8	1.99
3.26 to 3.50	7	1.36
3.01 to 3.25	6	.59
2.76 to 3.00	5	0
2.51 to 2.75	4	- .63
2.26 to 2.50	3	-1.30
2.01 to 2.25	2	-1.99
2.00 or less	1	-2.67

Table A-3
Normal Ogive Parameters for Items of the
Peaked Conventional Test

Item Reference Number	Item Discrimination (*a*)	Item Difficulty (*b*)
322	.81	-.97
640	.96	-.93
103	1.48	-.88
85	1.16	-.60
546	.63	-.73
671	.71	-.71
109	1.50	-.63
43	1.54	-.76
185	.79	-.62
123	.97	-.51
293	.76	-.52
149	.97	-.42
222	.74	-.46
33	.92	-.36
203	.94	-.36
313	1.00	-.34
46	.96	-.34
588	.51	-.43
58	.65	-.35
307	.78	-.30
155	.44	-.53
221	.92	-.26
37	.96	-.22
211	.86	-.22
143	1.17	-.14
156	.92	-.16
203	.81	-.17
154	.95	-.12
503	1.18	-.09
365	.95	-.10
234	.69	-.13
382	.99	-.01
157	.41	-.23
380	1.90	.11
292	.64	.01
176	.44	-.10
104	.91	.05
626	.93	.17
670	.88	.20
599	1.64	.16
205	.63	-.02
597	.65	0
283	1.75	.53
645	.68	.24
50	.68	.32
622	.45	.20
369	.78	.30
252	.42	.48
116	.49	.33
295	.63	.56
Mean	.89	-.20
S.D.	.34	.38

Table A-4
Normal Ogive Parameters for
Post-Test Items

Item Reference Number	Item Discrimination (a)	Item Difficulty (b)
2031	.83	-2.85
2151	.91	-2.75
2091	.68	-2.52
2152	.74	-2.46
2214	.77	-2.27
2219	.61	-2.08
2041	.74	-1.72
2094	.74	-1.72
2218	1.35	-1.57
2212	1.29	-1.39
2224	1.28	-1.17
2161	.99	- .99
2033	1.04	-1.01
2213	.74	- .92
2229	.71	- .88
2164	1.51	- .43
2216	1.59	- .42
2238	.93	- .41
2232	.88	- .39
2095	1.08	- .29
2220	1.03	- .27
2225	1.08	- .06
2234	1.09	- .03
2107	.91	.00
2043	.90	- .02
2053	1.34	.23
2237	1.31	.24
2046	1.86	.38
2050	1.86	.52
2172	1.19	.64
2111	1.13	.64
2049	1.28	.69
2233	1.36	.70
2042	1.47	.82
2055	1.52	.90
2115	1.70	1.05
2054	1.26	1.15
2239	1.74	1.33
2058	1.76	1.38
2179	1.88	1.47
2059	1.04	1.66
2119	1.94	1.89
2052	1.86	1.91
2060	1.86	2.62
Mean	1.22	- .19
S.D.	.40	1.37

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

1 ONR Branch Office
495 Summer Street
Boston, MA 02210
ATTN: Dr. James Lester

1 ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: Dr. Eugene Gloye

1 ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: Dr. Charles E. Davis

1 Dr. M. A. Bertin, Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco 96503

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Dr. H. Wallace Sinaiko
c/o Office of Naval Research
Code 450
Arlington, VA 22217

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Technical Director
Navy Personnel Research
and Development Center
San Diego, CA 92152

1 Assistant Deputy Chief of Naval
Personnel for Retention Analysis
and Coordination (Pers 12)
Room 2403, Arlington Annex
Washington, DC 20370

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Dr. Lee Miller
Naval Air Systems Command
AIR-413E
Washington, DC 20361

1 CDR Paul D. Nelson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Commanding Officer
Naval Health Research Center
San Diego, CA 92152
ATTN: Library

1 Chairman
Behavioral Science Department
Naval Command & Management Division
U.S. Naval Academy
Annapolis, MD 21402

1 Chief of Naval Education & Training
Naval Air Station
Pensacola, FL 32508
ATTN: CAPT Bruce Stone, USN

1 Mr. Arnold I. Rubinstein
Human Resources Program Manager
Naval Material Command (0344)
Room 1044, Crystal Plaza #5
Washington, DC 20360

1 Dr. Jack R. Borsting
U. S. Naval Postgraduate School
Department of Operations Research
Monterey, CA 93940

1 Director, Navy Occupational Task
Analysis Program (NOTAP)
Navy Personnel Program Support
Activity
Building 1304, Bolling AFB
Washington, DC 20336

1 Office of Civilian Manpower Management
Code 64
Washington, DC 20390
ATTN: Dr. Richard J. Niehaus

1 Office of Civilian Manpower Management
Code 263
Washington, DC 20390

1 Chief of Naval Reserve
Code 3055
New Orleans, LA 70146

1 Assistant to the Assistant Deputy Chief
of Naval Operations (Manpower)
Head, NAMPS Project Office
Room 1606, Arlington Annex
Washington, DC 20370
ATTN: Dr. Harry M. West

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Mr. George N. Graine
Naval Sea Systems Command
SEA 047C12
Washington, DC 20362

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. Norman J. Kerr

1 Principal Civilian Advisor
for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508
ATTN: Dr. William L. Maloy

1 Director
Training Analysis & Evaluation Group
Code N-00t
Department of the Navy
Orlando, FL 32813
ATTN: Dr. Alfred F. Smode

1 Chief of Naval Training Support
Code N-21
Building 45
Naval Air Station
Pensacola, FL 32508

1 Naval Undersea Center
Code 6032
San Diego, CA 92132
ATTN: W. Gary Thomson

1 LCDR C. F. Logan, USN
F-14 Management System
COMFITAEWWINGPAC
NAS Miramar, CA 92145

1 Navy Personnel Research
and Development Center
Code 01
San Diego, CA 92152

5 Navy Personnel Research
and Development Center
Code 02
San Diego, CA 92152
ATTN: A. A. Sjoholm

2 Navy Personnel Research
and Development Center
Code 310
San Diego, CA 92152
ATTN: Dr. Martin F. Wiskoff

2 Navy Personnel Research
and Development Center
Code 309
San Diego, CA 92152
ATTN: Mr. R. P. Thorpe

1 Navy Personnel Research
and Development Center
San Diego, CA 92152
ATTN: Library

1 Navy Personnel Research
and Development Center
Code 9041
San Diego, CA 92152
ATTN: Dr. J. D. Fletcher

1 D. M. Gragg, CAPT, MC, USN
Head, Educational Programs Development
Department
Naval Health Sciences Education and
Training Command
Bethesda, MD 20014

1 Mr. Victor H. Brown, Director
Career Training Analysis Group
Chief of Naval Education & Training
(Code N54)
Naval Air Station
Pensacola, FL 32507

Army

1 Technical Director
U.S. Army Research Institute for the
Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Headquarters
U.S. Army Administration Center
Personnel Administration Combat
Development Activity
ATCP-HRQ
Ft. Benjamin Harrison, IN 46249

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 Commandant
U.S. Army Infantry School
Fort Benning, GA 31905
ATTN: ATSH-DET

1 Deputy Commander
U.S. Army Institute of Administration
Fort Benjamin Harrison, IN 46216
ATTN: EA

1 Dr. Stanley L. Cohen
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Ralph Dusek
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Joseph Ward
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 HQ USAREUR & 7th Army
ODCSOPS
USAREUR Director of GED
APO New York 09403

1 ARI Field Unit - Leavenworth
Post Office Box 3122
Fort Leavenworth, KS 66027

1 Dr. James L. Raney
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Milton S. Katz, Chief
Individual Training & Performance Evaluation
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Air Force

1 Research Branch
AF/DPMYAR
Randolph AFB, TX 78148

1 Dr. G. A. Eckstrand (AFHRL/AST)
Wright-Patterson AFB
Ohio 45433

1 AFHRL/DOJN
Stop #63
Lackland AFB, TX 78236

1 Dr. Martin Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Instructional Technology Branch
AF Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Alfred R. Fregly
AFOSR/NL
[illegible]00 Wilson Boulevard
[illegible] VA 22209

1 AFHRL/PED
Stop #63
Lackland AFB, TX 78236

1 Major Wayne S. Sellman
Chief of Personnel Testing
HQ USAF/DPMYP
Randolph AFB, TX 78148

Marine Corps

1 Director, Office of Manpower Utilization
Headquarters, Marine Corps (Code MPU)
MCB (Building 2009)
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code RD-1)
Headquarters, U.S. Marine Corps
Washington, DC 20380

1 Chief, Academic Department
Education Center
Marine Corps Development and Education Command
Marine Corps Base
Quantico, VA 22134

1 Mr. E. A. Dover
2711 South Veitch Street
Arlington, VA 22206

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (G-P-1/62)
U.S. Coast Guard Headquarters
Washington, DC 20590

Other DOD

1 Military Assistant for Human Resources
Office of the Secretary of Defense
Room 3D129, Pentagon
Washington, DC 20301

1 Advanced Research Projects Agency
Administrative Services
1400 Wilson Boulevard
Arlington, VA 22209
ATTN: Ardella Holloway

1 Dr. Harold F. O'Neil, Jr.
Advanced Research Projects Agency
Human Resources Research Office
1400 Wilson Boulevard
Arlington, VA 22209

1 Dr. Robert Young
Advanced Research Projects Agency
Human Resources Research Office
1400 Wilson Boulevard
Arlington, VA 22209

1 Mr. Frederick W. Suffa
Chief, Recruiting and Retention Evaluation
Office of the Assistant Secretary of Defense, M&RA)
Room 3D970, Pentagon
Washington, DC 20301

12 Defense Documentation Center
Cameron Station, Building 5
Alexandria, VA 22314
ATTN: TC

Other Government

1 Dr. Lorraine D. Eyde
Personnel Research and Development Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. William Gorham, Director
Personnel Research and Development Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Vern Urry
Personnel Research and Development Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Harold T. Yahr
Personnel Research and Development Center
U.S. Civil Service Commission
1900 E Street, N.W.
Washington, DC 20415

1 Dr. Andrew R. Molnar
Technological Innovations in Education Group
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

1 U.S. Civil Service Commission
Federal Office Building
Chicago Regional Staff Division
Regional Psychologist
230 South Dearborn Street
Chicago, IL 60604
ATTN: C. S. Winiewicz

1 Dr. Carl Frederiksen
Learning Division, Basic Skills Group
National Institute of Education
1200 19th Street, N.W.
Washington, DC 20208

Miscellaneous

1 Dr. Scarvia B. Anderson
Educational Testing Service
17 Executive Park Drive, N.E.
Atlanta, GA 30329

1 Mr. Samuel Ball
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gerald V. Barrett
University of Akron
Department of Psychology
Akron, OH 44325

1 Dr. Bernard M. Bass
University of Rochester
Graduate School of Management
Rochester, NY 14627

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. A. Charnes
BEB 512
University of Texas
Austin, TX 78712

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
University of Southern California
Department of Psychology
University Park
Los Angeles, CA 90007

1 Dean W. W. Cooper
Carnegie-Mellon University
School of Urban and Public Affairs
Pittsburgh, PA 15213

1 Dr. H. Peter Dachler
University of Maryland
Department of Psychology
College Park, MD 20742

1 Dr. Rene' V. Dawis
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Norman R. Dixon
200 South Craig Street
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Marvin D. Dunnette
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Victor Fields
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Visiting Professor
University of California
Graduate School of Administration
Irvine, CA 92664

1 Dr. Robert Glaser, Co-Director
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Mr. Harry H. Harman
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
5640 Nicholson Lane
Rockville, MD 20852

1 Dr. M. D. Havron
Human Sciences Research, Inc.,
7710 Old Spring House Road
West Gate Industrial Park
McLean, VA 22101

1 HumRRO Central Division
400 Plaza Building
Pace Boulevard at Fairfield Drive
Pensacola, FL 32505

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
ATTN: Library

1 HumRRO Central Division/Columbus Office
Suite 23, 2601 Cross Country Drive
Columbus, GA 31906

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
ATTN: Dr. Robert Vineberg

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
2001 S Street, N.W., Suite 502
Washington, DC 20009

1 Dr. Roger A. Kaufman
203 Dodd Hall
Florida State University
Tallahassee, FL 32306

1 Dr. David Klahr
Carnegie-Mellon University
Department of Psychology
Pittsburgh, PA 15213

1 Dr. Alma E. Lantz
University of Denver
Denver Research Institute
Industrial Economics Division
Denver, CO 80210

1 Mr. W. E. Lassiter
Data Solutions Corporation
6849 Old Dominion Drive, Suite 211
McLean, VA 22101

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Corton Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Dr. William C. Mann
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, CA 90291

1 Mr. Edmond Marks
315 Old Main
Pennsylvania State University
University Park, PA 16802

1 Dr. Leo Munday, Vice President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. A. J. Pesch, President
Eclectech Associates, Inc.
P.O. Box 178
North Stonington, CT 06359

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Steven M. Pine
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
University of Southern California
Behavioral Technology Laboratories
3717 South Grand
Los Angeles, CA 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Montgomery College
Department of Psychology
Rockville, MD 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. Benjamin Schneider
University of Maryland
Department of Psychology
College Park, MD 20742

1 Dr. Arthur I. Siegel
Applied Psychological Services
404 East Lancaster Avenue
Wayne, PA 19087

1 Dr. Henry P. Sims, Jr.
Room 630 - Business
Indiana University
Bloomington, IN 47401

1 Dr. C. Harold Stone
1428 Virginia Avenue
Glendale, CA 91202

1 Mr. Dennis J. Sullivan
c/o HAISC, Building 119, M.S. 2
P.O. Box 90515
Los Angeles, CA 90009

1 Dr. K. W. Uncapher
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey, CA 90291

1 Mr. George Wheaton
American Institutes for Research
3301 New Mexico Avenue, N.W.
Washington, DC 20016

1 Dr. K. Wescourt
Stanford University
Institute for Mathematical Studies in the Social Sciences
Stanford CA 94305

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

1 Richard T. Mowday
College of Business Administration
University of Nebraska, Lincoln
Lincoln, NE 68588

1 Dr. John J. Collins
Vice President
Essex Corporation
6305 Caminito Estrellado
San Diego, CA 92120

1 Dr. Lyle Schoenfeldt
Department of Psychology
University of Georgia
Athens, Georgia 30602

1 Dr. Patrick Suppes, Director
Institute for Mathematical Studies
in the Social Sciences
Stanford University
Stanford, CA 94305

Previous Reports in this Series

73-1. Weiss, D.J. & Betz, N.E. Ability Measurement: Conventional or Adaptive? February 1973. (AD 757788).

73-2. Bejar, I.I. & Weiss, D.J. Comparison of Four Empirical Differential Item Scoring Procedures. August 1973.

73-3. Weiss, D.J. The Stratified Adaptive Computerized Ability Test. September 1973. (AD 768376).

73-4. Betz, N.E. & Weiss, D.J. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973. (AD 768993).

74-1. DeWitt, L.J. & Weiss, D.J. A Computer Software System for Adaptive Ability Measurement. January 1974. (AD 773961).

74-2. McBride, J.R. & Weiss, D.J. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (AD 781894).

74-3. Larkin, K.C. & Weiss, D.J. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (AD 783553).

74-4. Betz, N.E. & Weiss, D.J. Simulation Studies of Two-Stage Ability Testing. October 1974. (AD A001230).

74-5. Weiss, D.J. Strategies of Adaptive Ability Measurement. December 1974. (AD A004270).

75-1. Larkin, K.C. & Weiss, D.J. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (AD A006733).

75-2. McBride, J.R. & Weiss, D.J. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975. (AD A007572).

75-3. Betz, N.E. & Weiss, D.J. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (AD A013185).

75-4. Vale, C.D. & Weiss, D.J. A Study of Computer-Administered Stradaptive Ability Testing. October 1975. (AD A018758).

75-5. Weiss, D. J. (Ed.). Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (AD A018675).

75-6. Vale, C.D. & Weiss, D.J. A Simulation Study of Stradaptive Ability Testing. December 1975. (AD A020961).

76-1. McBride, J.R. & Weiss, D.J. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (AD A022964).

76-2. Miller, T.W. & Weiss, D.J. Effects of Time-Limits on Test-Taking Behavior. April 1976. (AD A024422).

Weiss, D.J. Final Report: Computerized Ability Testing. April 1976. (AD A024516).

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.

Copies of these reports are available, while supplies last, from:

Psychometric Methods Program
Department of Psychology
University of Minnesota
Minneapolis, Minnesota 55455