Effects of Knowledge of Results and Varying Proportion Correct on Ability Test Performance and Psychological Variables

J. Stephen Prestwood
University of Minnesota

Cronbach (1970, p. 35) has described ability tests as "those that seek to measure the maximum performance of the subject." It has long been recognized that for testees to achieve maximum levels of performance, they must be sufficiently motivated. The provision of immediate feedback or "knowledge of results" (KR) during testing has often been looked upon as one method for increasing testee motivation. On-line computerized testing has made the provision of KR a relatively simple matter. The ease with which KR can be administered is an added advantage of computerized adaptive testing which lies beyond the purely psychometric benefits of such procedures.

To study the effects and possible benefits of computer-administered KR, Betz and Weiss (1976a,b) administered multiple-choice tests of verbal ability to college undergraduates at the University of Minnesota. The tests were administered either with or without KR after each item response. Their data showed higher testee performance, as measured by maximum likelihood ability estimates, for students in the KR condition. Perceptions of test difficulty were more accurate for testees receiving KR; and these students also exhibited higher levels of motivation, as assessed by post-test measurements. Betz and Weiss' data also indicated that students' reactions to the provision of KR became more favorable as the proportion of positive feedback increased.

Because KR increased testee performance and motivation, and because testees reacted more favorably to the provision of KR as the proportion of positive feedback increased, an analysis of the joint effects of KR provision and the proportion of positive feedback (test difficulty) was initiated.

Method

Procedure

Subjects. Participating in this study were 561 undergraduate students enrolled in an introductory psychology course at the University of Minnesota in the fall of 1975. All subjects were volunteers who received points towards their final course grade for participation in the experiment. Students were sequentially assigned to experimental conditions.

Test administration. The students were tested at individual cathode-ray terminals (CRTs) connected to a Hewlett-Packard computer system. Instructional screens explaining the operation of the CRTs preceded the actual testing, and a proctor was present in the testing room to provide assistance in the operation of equipment. During the test, items were presented on the CRT screen; students responded by typing in a number corresponding to the chosen alternative for each of the 50 five-alternative multiple-choice vocabulary items.

Independent variables. A three-way factorial design was employed in the study. One factor was immediate knowledge of results (KR). Testees in the KR condition were informed by the computer immediately after a response whether it was correct or incorrect. After an incorrect response they were also told which of the alternatives was correct. Testees in the no-KR condition received no feedback. Another factor was ability-test strategy. Testees received either a conventional peaked ability test or a stradaptive ability test (Weiss, 1973). The third factor was test difficulty or proportion of positive feedback.

Three conventional tests and three stradaptive test-administration procedures were designed so that testees, on the average, would answer approximately 40%, 60%, or 80% of the test items correctly. Level of difficulty--high, medium, or low--was inversely related to the proportion of positive feedback a testee received, whether that feedback was explicit as in the KR condition or subjective as in the no-KR condition.

Items were chosen for the three peaked conventional tests on the basis of their normal ogive difficulty and discrimination parameters so that students would, on the average, be expected to answer 40%, 60%, or 80% of the questions correctly. The stradaptive tests were designed by constructing a stratified item pool with items grouped into nine non-overlapping difficulty strata. The items within each stratum were then arranged in decreasing order of discrimination.

The stradaptive branching routine normally branches to a different stratum depending on whether or not the preceding item was answered correctly (Weiss, 1973). For this study the procedure was modified so that branching to a more difficult stratum occurred whenever the current overall proportion-correct score for an individual was greater than a target value; and branching to a less difficult stratum occurred whenever the current proportion-correct score for an individual fell below the target value. The specific target values employed were determined by simulations and were chosen so that the actual final proportion-correct scores would be approximately .40, .60, or .80, as appropriate for the high-, medium-, and low-difficulty tests.

Dependent variables. Both the ability-test performance and the psychological reactions of the testees were of interest in the present study. Testee performance was measured by maximum-likelihood scores computed for each testee by solving the likelihood equation for the three-parameter logistic model of Birnbaum (1968, p. 459). Proportion-correct scores were also computed in order to ascertain the accuracy with which the target test difficulties were obtained. These latter scores, however, were not

used as ability measures per se, since the stradaptive test-administration program was designed to yield an arbitrary proportion-correct for each testee, and testees received different items. Furthermore, differences in proportion-correct scores between testees were predetermined, to a degree, by the construction of the three different conventional tests which were administered to them.

The psychological reactions of testees to the task were determined using the responses of testees to rating-scale items administered following the test. The four scales constructed from these items measured a testee's perception of the test's difficulty; his/her level of anxiety during testing; his/her motivation to do well on the test; and, for each testee in the KR conditions, his/her reactions to the provision of explicit feedback.

The scales in this experiment were factor analytically derived and differed slightly from those employed in Betz and Weiss (1976b). The alternatives on the scale items were weighted so that items received equal weighting on the scale and so that increasing scores on the four scales corresponded to the following: increasing motivation, increasing anxiety, increasingly positive reactions to the provision of KR, and perceptions of increasing difficulty.

Results

Ability-Test Data

Proportion-correct measures. Table 1 shows the mean and standard deviation of the proportion-correct measures for each experimental condition. The table shows that, on the average, the tests in each condition achieved the appropriate target proportion-correct with a good degree of accuracy.

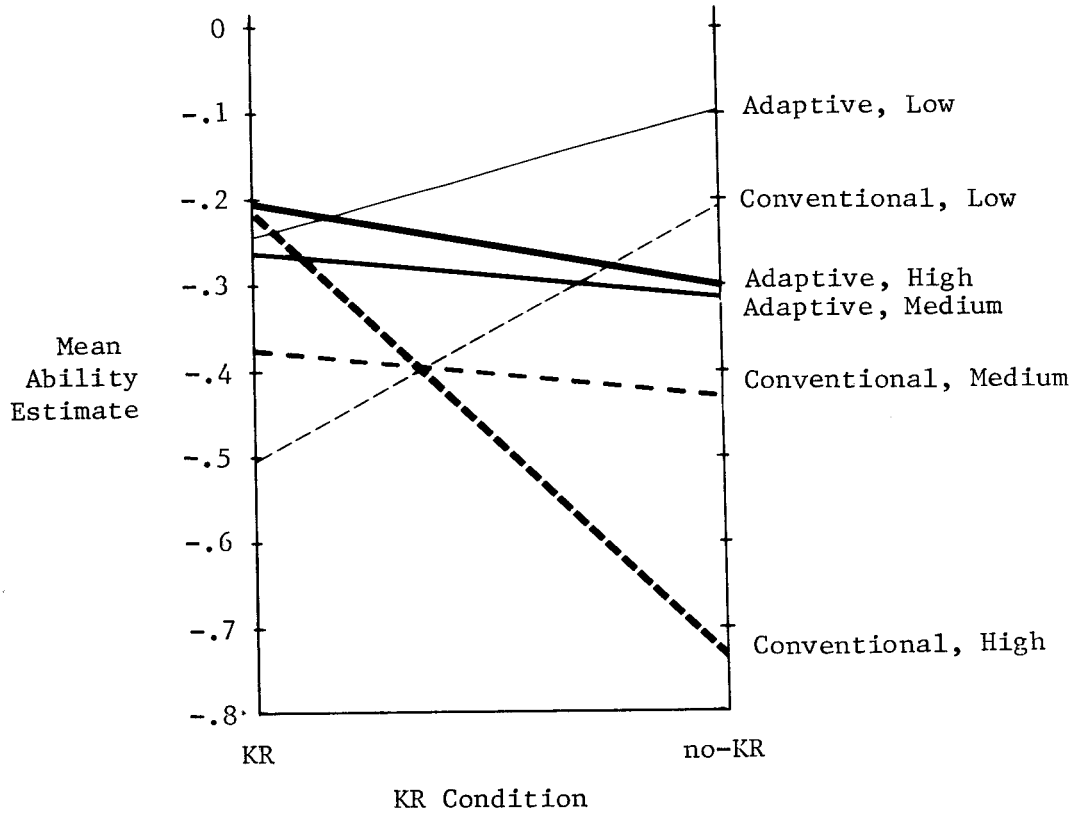
Table 1
Means and Standard Deviations of Proportion-Correct Scores for Conventional and Stradaptive Tests With and Without KR at Three Levels of Difficulty

Experimental Condition	Experimental Condition					
	KR			No-KR		
	N	Mean	S.D.	*N*	Mean	S.D.
Conventional Test						
Low Difficulty	48	.783	.106	45	.808	.103
Medium Difficulty	47	.608	.147	49	.592	.141
High Difficulty	46	.451	.188	46	.364	.153
Stradaptive Test						
Low Difficulty	44	.828	.064	45	.824	.046
Medium Difficulty	49	.617	.031	47	.610	.041
High Difficulty	49	.434	.103	46	.417	.076

The largest discrepancy (.051) was that between the target value of .400 and the actual value of .451 for the low-difficulty conventional test administered with feedback.

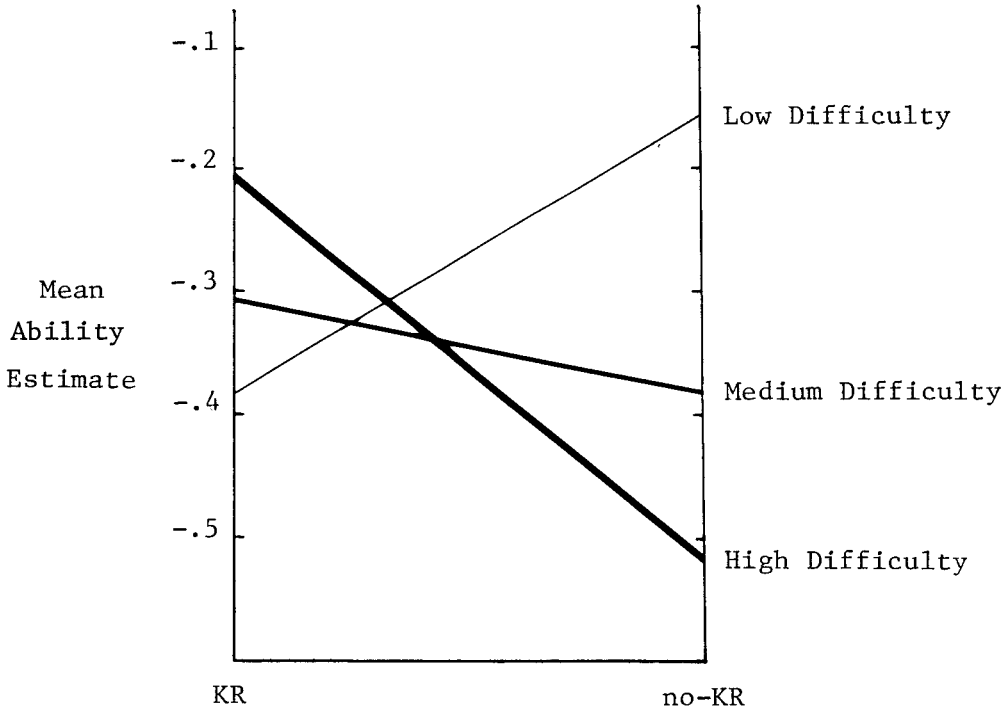
Maximum-likelihood ability estimates. Figure 1 shows the mean maximum-likelihood ability estimate for each of the 12 experimental conditions. The means for the KR conditions are plotted on the left vertical axis; the means for the conditions without KR are plotted on the right vertical axis. A three-way analysis of variance on the maximum-likelihood ability estimates showed a significant main effect for test strategy, $F(1, 549) = 3.984$, $p<.05$. The mean ability estimate for testees administered the stradaptive tests (-.239) was significantly higher than the mean for testees taking the conventional tests (-.415). There was also a marginally significant two-way interaction between the KR-provision and the test-difficulty factors, $F(2, 549) = 2.905$, $p\approx.054$. This interaction is shown graphically in Figure 2.

Figure 1
Mean Maximum-Likelihood Ability Estimates for
Adaptive and Conventional Tests of
Differing Difficulty Levels, by KR Condition



In Figure 2 the stradaptive and conventional test means have been combined. Interestingly, the effects of test difficulty on test performance were directionally opposite depending on whether or not KR was provided. When KR was provided, the mean testee ability estimate was highest on the most difficult tests (-.209) and lowest on the relatively easy tests (-.375). The mean ability estimates for testees in the no-KR conditions were highest on the easy tests (-.156) and lowest on the difficult tests (-.523).

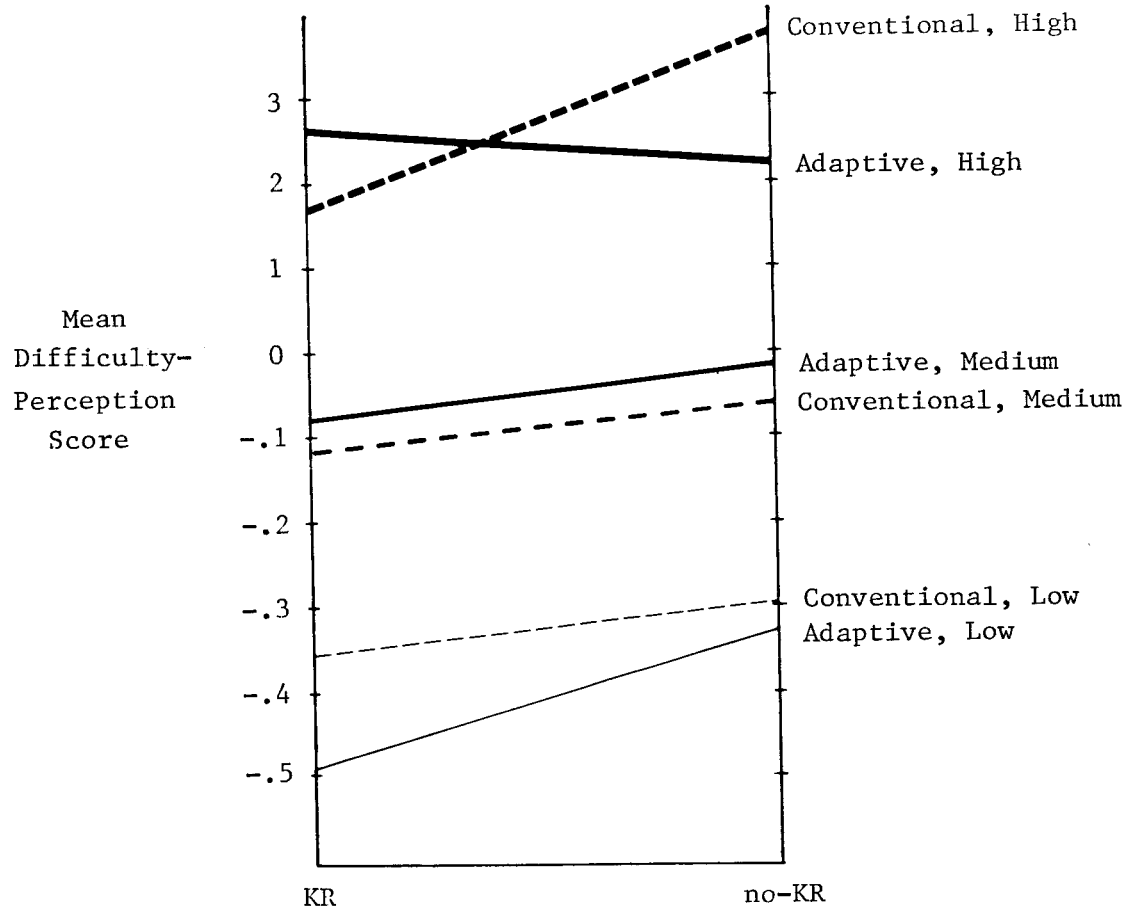
Figure 2
Mean Maximum-Likelihood Ability Estimates As
a Function of KR Condition and Test Difficulty



Psychological Reactions

Figure 3 shows the mean difficulty-perception scores as a function of experimental condition. A three-way analysis of variance assessing the effects of KR provision, test strategy, and test difficulty on the difficulty-perception data showed the expected main effect of test difficulty, $F(2, 549) = 163.243$, $p<.01$, and an additional main effect for the KR variable $F(1, 549) = 8.334$, $p<.01$. Tests administered without KR were perceived as significantly more difficult than tests administered with KR. There was also a marginally significant three-way interaction, $F(2, 549) = 2.810$, $p\simeq.059$, due to the fact that the adaptive tests administered without KR differed less in perceived difficulty than did the adaptive tests administered with KR.

Figure 3
Mean Difficulty-Perception Scores for
Adaptive and Conventional Tests of
Differing Levels, by KR Condition

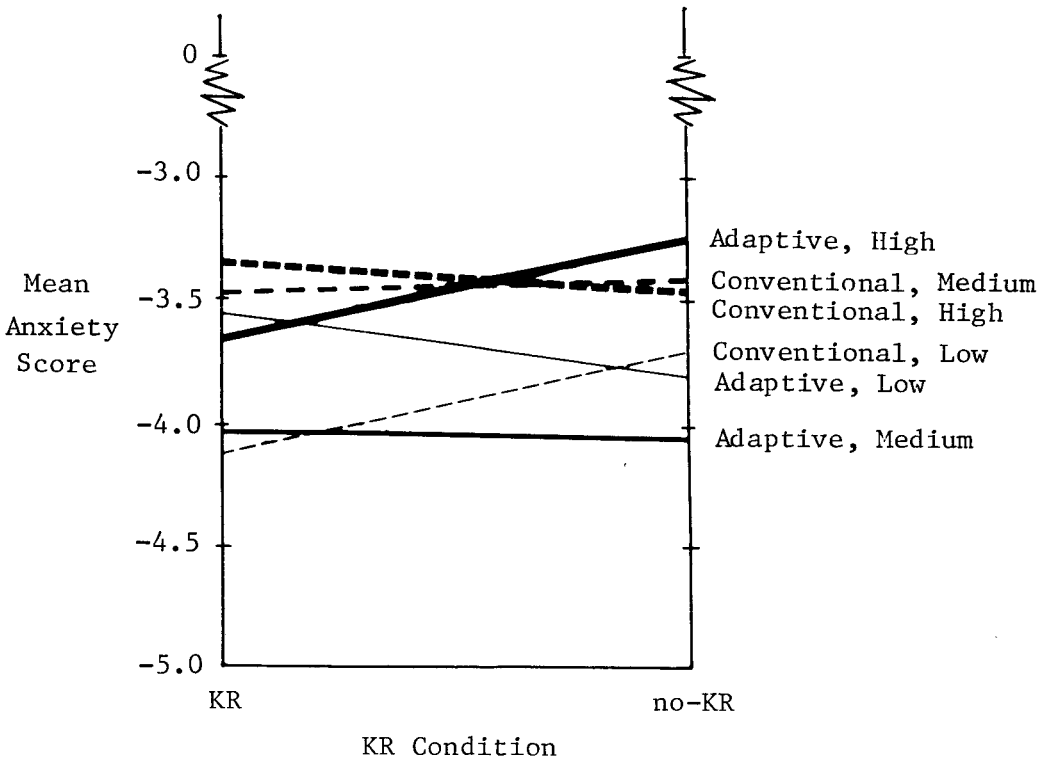


Mean anxiety scores are plotted as a function of experimental condition in Figure 4. An analysis of variance of these scores showed no effects of KR provision, test strategy, or test difficulty on mean level of anxiety.

Figure 5 shows the mean motivation scores for each of the 12 experimental conditions. A three-way analysis of variance of the motivation-scale data indicated a significant main effect for the KR factor, $F(1, 549) = 5.098$, $p<.05$. The mean motivation score for testees receiving KR (.096) was significantly higher than that for testees not receiving KR (-.110).

Figure 6 shows the mean KR-reaction score plotted as a function of test strategy for the high-, medium-, and low-difficulty tests. A two-way analysis

Figure 4
Mean Anxiety Scores for
Adaptive and Conventional Tests of
Differing Levels, by KR Condition



of variance assessing the effects of test strategy and test difficulty on testee reactions to KR showed no significant effects of experimental condition upon mean KR-reaction scores.

The endorsement frequencies of response-options on the KR-reaction questions are shown in Table 2. Of the 283 testees receiving immediate feedback, 87% felt that feedback made the test much more interesting; 86% felt that feedback did not interfere with their ability to concentrate on the test; 76% reported that feedback did not make them nervous; and 81% were very interested in knowing whether their answers were right or wrong. Ninety-two percent indicated that they liked getting feedback.

Table 3 shows the Pearson-product-moment correlation coefficients between the four psychological reactions scales. Those correlations involving the KR-reaction scale were based on the 283 students in the KR conditions. The other correlations were based on all 561 testees. As perceived difficulty increased, anxiety scores increased and motivation scores decreased. Reported anxiety was positively but not highly correlated with motivation scores. Students receiving KR reacted more favorably to KR as the perceived difficulty of the tests decreased.

Figure 5
Mean Motivation Scores for Adaptive and
Conventional Tests of Differing Difficulty Levels, by KR Condition

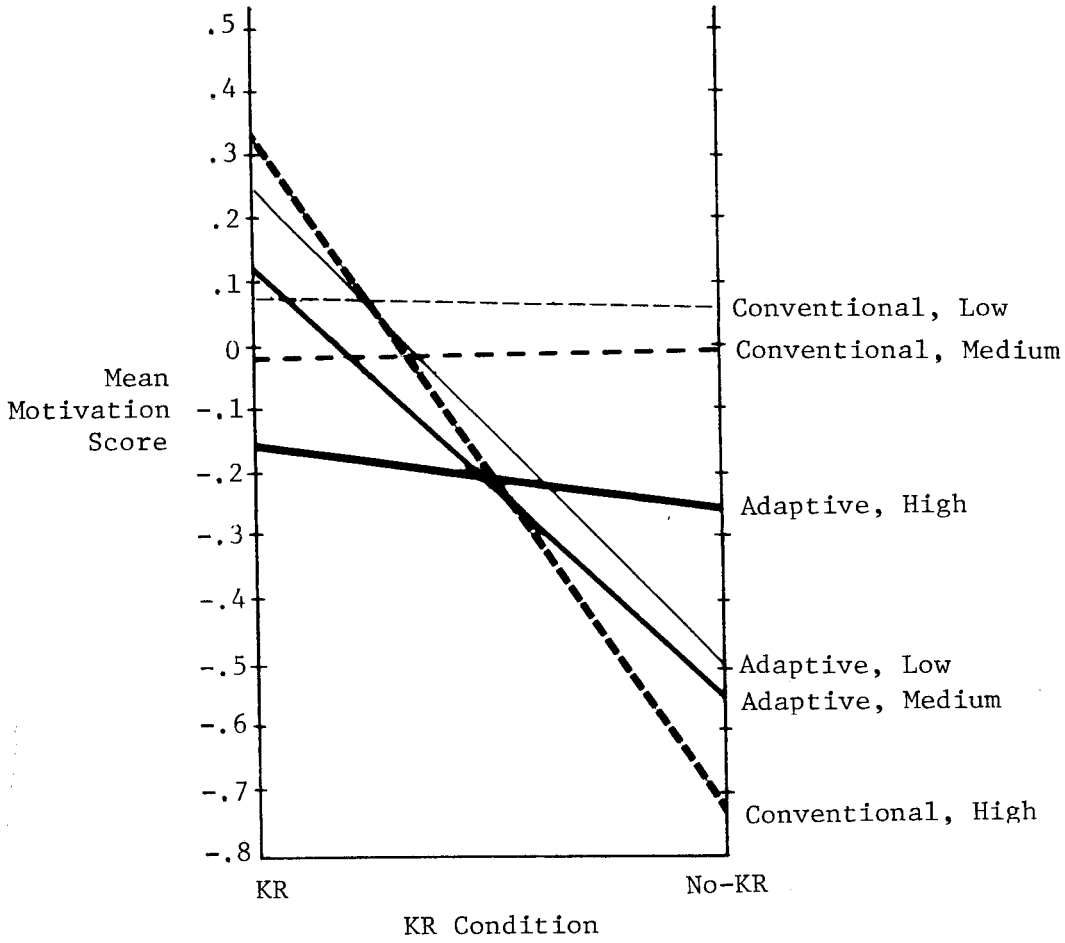


Figure 6
Mean KR-Reaction Scores for Adaptive and
Conventional Tests of Differing Difficulty Levels

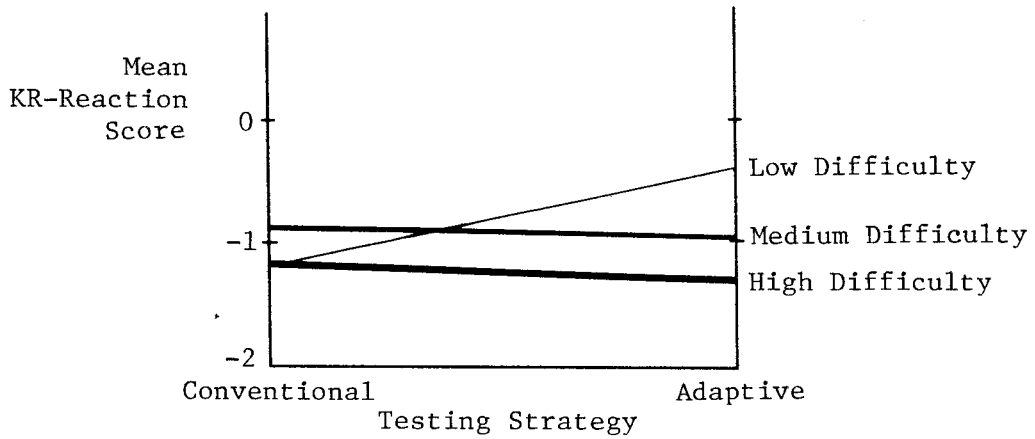


Table 2
Relative Frequencies of Response-Option Endorsement for KR-Reaction Items

Item	Endorsement Proportion
Did getting feedback on this test make it more interesting or less interesting?	
1. Much more interesting	.87
2. Somewhat more interesting	.11
3. Didn't make any difference	.00
4. Somewhat less interesting	.01
5. Much less interesting	.01
Did receiving feedback after each question interfere with your ability to concentrate on the test?	
1. No, not at all	.86
2. Yes, somewhat	.12
3. Yes, moderately so	.04
4. Yes, very much so	.02
Did getting feedback after each question make you nervous?	
1. No, not at all	.76
2. Yes, somewhat	.22
3. Yes, moderately so	.01
4. Yes, very much so	.01
Were you interested in knowing whether your answers were right or wrong?	
1. I was very interested	.81
2. I was moderately interested	.14
3. I was somewhat interested	.04
4. I didn't care at all	.00
How do you feel about getting feedback?	
1. I'd rather not know whether my answers were right or wrong	.07
2. I really don't care whether I get feedback or not	.01
3. I liked getting the feedback	.92

Table 3
Intercorrelations of Psychological-Reactions Scales

	Scale			
Scale	DIF	ANX	MOT	KR
DIF	-	.21**	-.17**	-.15*
ANX	.21**	-	.13**	-.16*
MOT	-.17**	.13**	-	.25**
KR	-.15*	-.16*	.25**	-

* Significant at the .05 level.
** Significant at the .01 level.

Discussion

The present data do not replicate the Betz and Weiss (1976a) finding that KR increased the average ability estimate of testees. Nor did these data find that anxiety was higher on the adaptive test--a finding reported by Betz and Weiss (1976b). The data in the present study showed higher ability estimates for students taking the adaptive test. Betz and Weiss (1976a) reported a similar effect,but only for a group of relatively lower ability students. There is agreement, however, between the present study and the Betz and Weiss (1976b) study,that the average motivation is higher for testees receiving KR. The increase in motivation accompanying the provision of KR may be partially due to the fact that testees receiving KR perceived the tests as being less difficult, on the average, than did testees not receiving KR. Also, it was noted that the scores on the difficulty-perception scale had a significant, although modest, negative correlation with motivation scores.

The marginally significant, but highly provocative, interaction of test difficulty and KR in the analysis of ability estimates indicates that the provision of KR may affect the performance of testees differentially,depending on the difficulty of the task. Although students reacted very favorably to KR regardless of the proportion of positive feedback, it would seem to be important that the effects of KR's provision *on test performance* be carefully investigated before it is provided under new sets of conditions.

References

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability-test performance (Research Report 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (NTIS No. ADA027147)(a)

Betz, N. E., & Weiss, D. 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. (NTIS No. AD027170) (b)

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental Test Scores. Reading, MA: Addison-Wesley, 1968, 532-552.

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

Weiss, D. J. The stratified adaptive computerized test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1973. (NTIS No. AD768376)

Acknowledgments

This research was supported by contract No. N00014-76-C-0243, NR150-382 with the Personnel and Training Research Programs of the Office of Naval Research, David J. Weiss Principal Investigator.