Knowledge of Results and the Proportion of Positive Feedback on Tests of Ability

J. Stephen Prestwood
University of Minnesota

Students were administered one of three conventional or one of three stradaptive vocabulary tests with or without knowledge of results (KR). The three tests of each type differed in the expected proportion of correct responses to the test items and thus in the expected proportion of positive feedback. Results indicated that the mean maximum-likelihood estimates of individuals' abilities varied as a joint function of KR-provision and test difficulty. Students receiving KR scored highest on the most-difficult test and lowest on the least-difficult test; students receiving no KR scored highest on the least-difficult test and did most poorly on the most-difficult test. Although the students perceived the differences in test difficulty, there were no effects on mean student anxiety or motivation scores attributable to difficulty or proportion of positive feedback alone. Regardless of the proportion of positive feedback, students reacted very favorably to receiving KR, and its provision increased the mean level of reported motivation.

Sufficient motivation is necessary in order to adequately measure an individual's ability (e.g., Cronbach, 1970; Samuda, 1975). The provision of immediate feedback or "knowledge of results" (KR) during testing is one possible method for increasing or maintaining high levels of motivation (Bayroff, 1964; Betz, 1975; Ferguson & Hsu, 1971; Strang & Rust, 1973; Zontine, Richards, & Strang, 1972). 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 testing which lies beyond the purely psychometric benefits of computerized adaptive procedures.

To study the effects and possible benefits of computer-administered KR, Betz and Weiss (1976a, 1976b) 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 conditions. The data also indicated that students' reactions to the provision of KR became more favorable as the proportion of *positive* feedback increased. Because, in the Betz and Weiss studies, KR increased performance and because individuals reacted more favorably to the provision of KR as the proportion of positive feedback increased, an analysis of the joint effects of provision of KR and the proportion of positive feedback (test difficulty) was initiated.

Method

Procedure

Subjects. Participating in this study were 561 undergraduate students enrolled in an intro-

ductory psychology course at the University of Minnesota in the fall of 1975. All students were volunteers who received points towards their final course grade for participation in the experiment. Students were sequentially assigned to experimental conditions.

Test administration. All students were tested at individual cathode-ray terminals (CRTs) connected to a Hewlett-Packard 9600E Real-Time computer system. Instructional screens explaining the operation of the CRTs preceded the actual testing. Each test consisted of 50 five-alternative multiple-choice vocabulary items. During the test, items were presented on the CRT screen and students responded by typing in a number corresponding to the chosen alternative for each item. Students were informed that they would have as much time as they needed to finish the test. Following the test, the students responded to a series of questions designed to assess their psychological reactions to the test.

Design

Independent variables. A three-way factorial design was employed in the study. One factor was immediate knowledge of results (KR). Students in the KR condition were informed by the computer immediately after their response to an item whether the response was correct or incorrect. After an incorrect response, they were told which of the alternatives was correct. Students in the no-KR condition did not receive this feedback. Ability-test strategy was a second factor. Individuals received either a conventional peaked ability test or a stradaptive ability test (Weiss, 1973). A third factor was test difficulty or proportion of positive feedback. Fifty items were chosen for each of the conventional tests so that the mean normal-ogive discriminations would be equal (approximately $a = .80$) on the three tests and so that students, on the average, would answer approximately 40%, 60%, or

80% of the test questions correctly. For the adaptive tests the usual stradaptive administration procedure was modified in this study so that branching to a more (or less) difficult item was based on whether the individual's overall proportion correct was over (or under) a target value. In order to achieve average proportion-correct scores of .40, .60, and .80 for the low-, medium-, and high-difficulty tests, target values of .36, .60, and .84 were employed for the three tests, respectively. The target values were determined from a prior analysis of simulated responding. Branching was terminated after the administration of 50 items. Level of difficulty (high, medium, or low) was inversely related to the proportion of positive feedback an individual received, whether that feedback was explicit, as in the KR condition, or subjective, as in the no-KR condition.

Dependent variables. Both the ability-test performance and the psychological reactions of the testees were of interest. Performance was measured by maximum-likelihood ability estimates computed for each individual by solving the likelihood equation for Birnbaum's (1968, p. 459) three-parameter logistic model. Proportion-correct scores were also computed in order to assure that the tests were of appropriate difficulty. Proportion correct, however, was not used as an ability measure per se, since the stradaptive test-administration procedure was designed to yield an arbitrary proportion correct for each student and since differences in proportion-correct scores between students administered the three different conventional tests were determined to some degree by the construction of those tests.

The reactions of individuals to the tests were determined using the responses of the students to psychological reactions items administered following the test. Four scales constructed from these items measured students' perceptions of the test's difficulty; their level of anxiety during testing; their motivation to do well on the test; and, for testees in the KR conditions, their reactions to the provision of KR.

Results

Ability Test Data

The mean proportion-correct measures for each condition showed that tests in each condition achieved the appropriate target proportion correct with a good degree of accuracy. The largest discrepancy (.051) occurred between the target value of .400 and the actual value of .451 for the low-difficulty conventional test administered with KR.

Table 1 shows the means and standard deviations of the maximum-likelihood ability estimates for students in each of the experimental conditions. An analysis of variance of these estimates revealed a significant main effect for test strategy, $F(1,549) = 3.98$, $p < .05$. The mean ability estimate for students administered the stradaptive tests (.24) was significantly higher than the mean for students administered the conventional tests (−.42).

There was also a marginally significant two-way interaction between the provision of KR and test-difficulty factors, $F(2,549) = 2.91$, $p \cong .054$. This interaction is shown graphically in Figure 1. Interestingly, Figure 1 shows that the effects of test difficulty on test performance were opposite in direction, depending on whether or not KR was provided. When KR was provided, the mean student ability estimate was highest on the most-difficult tests (−.21) and lowest on the least-difficult tests (−.38). The mean ability estimate for students in the no-KR conditions was highest on the least-difficult tests (−.16) and lowest on the most-difficult tests (−.52).
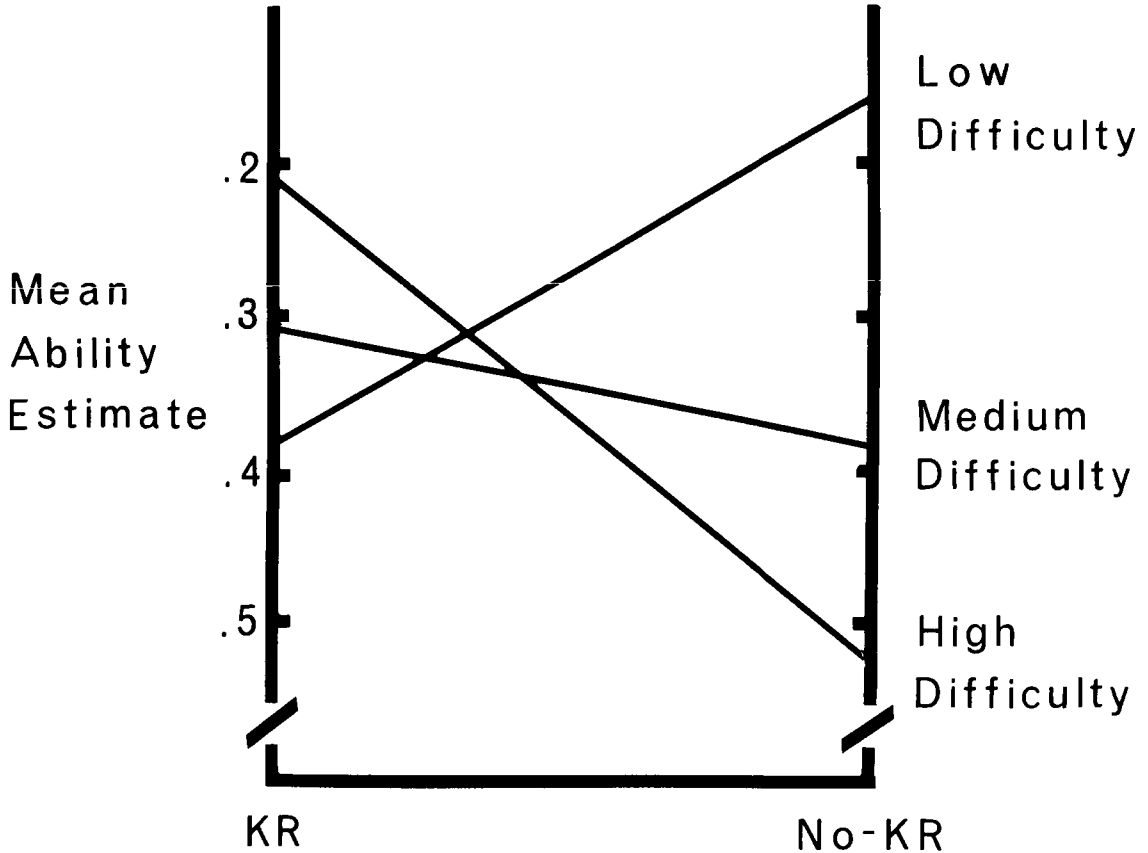
Psychological Reactions to Testing

An analysis of variance of the students' perceptions of test difficulty showed that the expected main effect due to test difficulty was

Table 1

Means and Standard Deviations of Maximum Likelihood Ability
Estimates for Conventional and Adaptive Tests With and Without
KR at Three Levels of Difficulty

Experimental Condition	Experimental Condition						Combined Conditions		
	KR			No-KR			Combined Conditions		
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.
Conventional Test									
Low Difficulty	48	−.50	.98	45	−.21	1.00	93	−.36	.99
Medium Difficulty	47	−.38	1.15	49	−.44	.96	96	−.41	1.06
High Difficulty	46	−.21	1.28	46	−.74	1.23	92	−.48	1.28
Adaptive Test									
Low Difficulty	44	−.24	1.05	45	−.10	.93	89	−.17	.99
Medium Difficulty	49	−.26	.81	47	−.32	.99	96	−.29	.90
High Difficulty	49	−.20	1.13	46	−.31	.98	95	−.26	1.05
Combined Conditions									
Conventional Test	141	−.37	1.14	140	−.46	1.08	281	−.42	1.11
Adaptive Test	142	−.23	.99	138	−.25	.97	280	−.24	.98
Low Difficulty	92	−.38	1.02	90	−.16	.96	182	−.27	.99
Medium Difficulty	96	−.32	.99	96	−.38	.97	192	−.35	.98
High Difficulty	95	−.21	1.20	92	−.52	1.13	187	−.36	1.17
Total	283	−.30	1.07	278	−.36	1.03	561	−.33	1.05

Figure 1
Mean Ability Estimate as a Function
of KR and Test Difficulty

highly significant, $F(2,549) = 163.24$, $p < .001$. In addition, the main effect of KR-provision was also significant, $F(1,549) = 8.33$, $p < .01$. Students receiving KR perceived the tests as being less difficult than did those not receiving KR.

A three-way analysis of variance of student's reported motivation indicated a significant main effect for the KR factor, $F(1,549) = 5.10$, $p < .05$. The mean Motivation score for individuals receiving KR (.10) was significantly higher than that for individuals not receiving KR (−.32).

Of the 283 students receiving KR, 87% indicated on the post-test questions that KR made

the test "much more interesting," 81% were very interested in knowing whether their answers were right or wrong, and 92% indicated that they liked getting immediate feedback.

Discussion

Betz and Weiss (1976a) found that the provision of KR increased the average maximum likelihood ability estimate of students administered a stradaptive or a conventional peaked test. They also found that KR increased the average number-correct score on the conventional test. The present data did not replicate these find-

ings. In this study the average ability estimate of students in the KR conditions and the average of students in the no-KR conditions did not differ significantly; nor did the present data show a higher level of anxiety on the adaptive test—a finding reported in Betz and Weiss (1976b).

In this study, the mean maximum likelihood ability estimate was higher on the adaptive test than on the conventional test. In addition, the data suggest that average motivation is increased by the provision of KR. This latter effect may be due to the fact that students not provided KR rated the tests as more difficult than did students receiving KR.

Prestwood and Weiss (1977) analyzed student perceptions of individual item difficulties and showed that items with normal-ogive difficulties somewhat below the estimated ability levels of the students were, on the average, perceived as being "about right" in difficulty. They suggested that by tailoring tests to ensure that the item difficulties are psychometrically optimal, adaptive strategies may also be tailoring tests so that, in effect, all of the items are perceived by testees as being too difficult. Their conclusion was that the psychological effects of such a procedure should be investigated more fully. The data reported in the present study show no significant effects on mean level of anxiety or motivation due to test difficulty alone. These data suggest that although students perceive items below their level of ability as being most appropriate in difficulty, these perceptions do not adversely affect motivation when test difficulty is in the range employed in this study (proportion correct of .40 to .80).

The marginally significant, but highly provocative, interaction of test difficulty and KR-provision in the analysis of maximum-likelihood ability estimates indicates that the provision of KR may affect the performance of individuals differentially, depending on the difficulty of the task. On a conventional test, individual students may receive different tests in a psychological sense. Although each student will receive the same items, low-ability students may be re-

sponding to a relatively "difficult" test, whereas high-ability students may be responding to a relatively "easy" test. If the provision of KR has differential effects depending on the difficulty of the task, students of different ability levels may react differently to the provision of KR according to their perceptions of the task's difficulty, even though they are administered the same set of test items. On an adaptive test, where the difficulty of the task is modified so that each individual answers approximately the same proportion of items correctly, the differential effects of KR and test difficulty should be minimized. Although students reacted very favorably to KR regardless of the proportion of positive feedback they received, the possible interaction of test difficulty and KR-provision in the analysis of ability estimates suggests that the effects of KR's provision on *performance* should be carefully investigated before KR is provided under new sets of conditions.

References

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

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, November 1976. (NTIS No. AD A018675)

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, June 1976. (NTIS No. AD A027147) (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, June 1976. (NTIS No. AD A027170) (b)

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, MA: Addison-Wesley, 1968.

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

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

Prestwood, J. S., & Weiss, D. J. *Accuracy of perceived test-item difficulties* (Research Report 77–3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, May 1977. (NTIS No. AD A035393)

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

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.

Weiss, D. J. *The stratified adaptive computerized ability test* (Research Report 73–3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376)

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.

Acknowledgments

Author's Address

Send requests for reprints and further information to J. Stephen Prestwood, Psychometric Methods Program, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455.