DOCUMENT RESUME

ED 102 276 UD 014 864

AUTHOR Reilly, Richard R.; Evans, Franklin R.
TITLE The Effects of Test Time Limits on Performance of Culturally Defined Groups.
PUB DATE Sep 74
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, Louisiana, September 1974)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Academic Aptitude; Aptitude Tests; College Students; Group Tests; *Minority Groups; Objective Tests; *Performance Factors; Response Style (Tests); Social Differences; *Standardized Tests; Statistical Analysis; *Testing Problems; Test Interpretation; Test Wiseness; *Timed Tests

ABSTRACT

One of the many criticisms leveled at standardized testing is that the time limits commonly used require a speed component of performance which may act to the disadvantaged of certain culturally defined groups. Recent studies by the authors examined the question of differential time limits and group performance for standardized academic aptitude tests. Both studies deleted items from an experimental section in order to manipulate the speed with which an examinee was required to work. The present study sought to extend the range of culturally defined groups and presented an opportunity to use a more direct method of manipulating working speed. A 27 item reading comprehension section of a national academic aptitude test was administered under two conditions. Under one condition the section was administered with the usual 30 minute time limit and under the other condition the time limit was extended to 40 minutes. Data for the study were collected in October and December of 1971. The majority of subjects were in their senior year in college. The categories of group membership included black, white, Chicano, and Oriental; male or female. As in the first two studies, the results rejected the hypothesis that allowing more liberal time limits would be more beneficial to minority groups than to the majority group. (Author/JM)

The Effects of Test Time Limits on Performance of Culturally Defined Groups

Richard R. Reilly and Franklin R. Evans
Educational Testing Service

One of the many criticisms leveled at standardized testing is that the time limits commonly used require a speed component of performance which may act to the disadvantage of certain culturally defined groups. Since the stated intent of a most standardized academic aptitude tests is to measure power without placing undue emphasis on speed, an important consideration for any test is the extent to which working speed is related to performance and whether manipulating the usual test time limits would result in differential levels of performance for different categories of examinees. Recent studies by Evans and Reilly (1972, 1973) examined the question of differential time limits and group performance for standardized academic aptitude tests. Both studies were limited, however, in the range of cultural groups included and the methodology employed. Because of the standardization required for nationally administered academic aptitude tests, both studies did not actually change the time limits but deleted items from an experimental section in order to manipulate the speed with which an examinee was required to work. Analyses were then carried out on a set of items that were common to different rate of work conditions. The present study sought to extend the range of culturally defined groups and, in addition, presented an opportunity to use a more direct method of manipulating working speed.

Methods

A 27 item reading comprehension section of a national academic aptitude test was administered under two conditions. Under condition A the section

was administered with the usual 30 minute time limit and under condition B the time limit was extended to 40 minutes. This 1/3 increase in time limit was felt to be sufficient to allow most examinees to finish regardless of personal characteristics. The two conditions were randomly assigned to test center. Assignment of condition to test center rather than subject was necessary because of the administrative problems which different time limits within the same center would create.

Data for the study were collected in October and December of 1971 during regular national administrations. The majority of subjects were in their senior year in college. The categories of group membership which were determined by responses to two questions on the last page of the test booklet included black male, black female, white male, white female, Chicano, and Oriental (because of the small proportion of females in these latter two groups, within-group sex breakdowns could not be done).

Two major analyses were performed to test the hypothesis that altering time limits would have no differential effect on the performance of the different categories of examinees. The first analysis was a two-way modified analysis of covariance with the six groups described above serving as levels of one factor and the two time conditions serving as levels of the other factor.

Since it was known that the experimental section would be highly correlated with the reading comprehension section in the standard part of the test, it was clear that under the proper experimental conditions the standard part of the test could serve as a covariate and reduce the experimental error. It was also clear that using a regular analysis of covariance was not appropriate because certain assumptions were not met, specifically non-random assignments to culturally defined groups.

Although a quasi-random procedure was employed for assigning test form to subjects within each group, there were rather larger differences in mean scores on the covariate across the six groups and there was no guarantee that the same regression line could be used to "adjust" scores for all six groups. In this study, however, primary interest was on the interaction of cultural group with time conditions, and since assignment within group was roughly random, a slightly different linear model used by the present authors in a previous study (Evans & Reilly, 1973) was used in an attempt to overcome some of the problems of traditional analysis of covariance. The following model was used:

$$Y_{ijk} = \alpha_i + \rho_i x_{ijk} + \beta_j + \gamma_{ij} + e_{ijk}$$

where x and y scores are expressed as deviations from their respective grand means; ρ_i and α_i represent within group slopes and intercepts, respectively, x_{ijk} is a covariate; β_j is the main effect due to time condition; γ_{ij} is the interaction between group and time condition; and e_{ijk} is the error term.

The model actually uses separate regression lines derived from data across treatments within a given cultural group, and in effect, the dependent variable becomes the deviation of the special section score about the within group regression line. With such a model main effects due to group differences could not be tested. In the present study, however, there was no interest in testing group differences, and the use of separate within group regression lines should have had the effect of merely reducing the error term to be used in testing the hypotheses of interest, that is, the main effects due to time condition, and especially the interaction effect.

A second analysis examined the effects of time limits on performance of blacks and whites and males and females in a 2^3 factorial design with race, sex and time condition as the three factors.

Insert Table 1 about here

Results and Discussion

The results of the first analysis were similar to those reported previously (Evans & Reilly, 1972) for blacks and whites as shown in Table 1. A slight but significant time condtion effect was observed but no significant interaction effect resulted. Thus, it appears that increasing the time allowed to complete the section had a slightly beneficial effect for all examinees but that this effect was not differential among the various groups identified. On the average, the mean difference in raw score between the more and less stringent time limits ranged from 4/5 to 8/5 of a point (see Table 2). It should also be mentioned that the number of degrees of freedom for the F tests in the present study were quite large such that even a very small proportion of variance accounted for by a given factor would have resulted in a significant F ratio. This time condition effect, in fact, accounted for less than one percent of the variance.

Insert Table 2 about here

The results of the 2^3 factorial analysis of variance revealed that none of the interaction effects were statistically significant. A significant F ratio was observed, however, for each of the main effects (race, time condition, and sex) but these were slight and did not account for much of the variance observed in the dependent variable scores. Of the three, the main effect due to race was the strongest and accounted for only 3 percent of the total variance.

Thus, in three studies by the authors (Evans & Reilly, 1972; 1973; and the present study) involving college seniors from various cultural groups (blacks, whites, Chicanos, etc.) experimental aptitude tests were administered to random portions of each group under varying time conditions to test the hypotheses that allowing more liberal time limits would be more beneficial to minority groups than to the majority group. The hypotheses was rejected in all three studies. Previous research conducted by others has led to conflicting results. For example, Dubin, Osburn, and Winick, 1969, working with high school students and a highly speeded test, found no interaction between time limit and race but Knapp, 1960, working with adult Mexican and American subjects and Moreton and Butcher, 1963, working with British school children (rural and urban) did find significant group by time condition interactions. These apparent contradictions are probably due to differences in the design of the various studies. The test sophistication of the various groups (high in the present authors' work and very low for Knapp's Mexicans and Moreton and Butcher's rural children) is certainly one important difference. Confounded with degree of test sophistication among the groups studied is the degree of speededness existing in the original test. The aptitude tests used by the present authors were not meant to be speeded while those used by Dubin, Osburn, and Winick were highly speeded tests on which one would expect all groups to benefit substantially from more liberal time limits.

A study being conducted by one of the present authors hopes to overcome some of these problems mentioned above. In that study rural and urban high school students from three culturally district groups (blacks, Chicanos, and whites) will be given both mathematics and verbal aptitude tests under varying time conditions.

Conclusions .

The present study replicated quite closely the findings of the previous study. While allowing examinees more time appeared to result in a very slight but significant increase in mean scores, the magnitude of those gains appeared to be relatively constant across various culturally defined groups. The apparent conflict of results with previous research is seen as due to differences in methodology. Further research is underway to clarify the findings.

References

Dubin, J. A., Osburn, H., & Winick, D. M. Speed and practice effects on Negro and white test performers. Journal of Applied Psychology, 1969, 53, 19-23.

Evans, F. R., & Reilly, R. R. A study of speededness as a source of test bias. Journal of Educational Measurement, 1973, 123-131.

Evans; F. R., & Reilly, R. R. A study of test speededness as a potential source of bias in the quantitative score of the admission test for graduate study in business. Research in Higher Education, 1973, 173-183.

Knapp, R. R. The effects of time limits on the intelligence test performance on Mexican and American subjects. Journal of Educational Psychology, 1963, 33, 22-30.

Moreton, C. A., & Butcher, H. J. Are rural children handicapped by the use of speeded tests in selection procedures? British Journal of Educational Psychology, 1963, 33, 22-30.

Table 1

Summary of Analysis of Covariance and Analysis of Variance

	Covariance Analysis			Analysis of Variance		
	R^2	df	F	R^2	df	F
Factor						
Group	.5693[1]	5/4610	1.29	.1012[1]	5/4616	31.31*
Time Condition	.5648	1/4610	54.66*	.1692	1/4616	12.30*
Group X Time Condition	.5690	5/4610	1.93	.1742	5/4616	.34
All Variables	.5699			.1750		

*$p < .01$

[1]Note.--The R^2 values are the squared multiple correlations, with the dummy variables representing each factor indicated removed.

Table 2

Differences in Group Mean Scores for 30 and 40 Minute Special Sections[1]

Group	Difference in Means
Black Male	1.34
Black Female	1.29
White Male	1.58
White Female	.81
Chicano	1.11
Oriental	1.60

[1]All differences were in favor of the 40 minute section.