nomics, future studies, operations research, political science, philosophy of science, psychology, statistics, and other pertinent subjects.

MARTIN SHUBIK
*Department of Administrative Sciences, Yale University, New Haven, Connecticut 06520*

## Test Bias

The letter (24 Dec. 1971, p. 1278) by Kenneth Clark and Lawrence Plotkin was meant to correct what they describe as "three egregious misstatements" in Stanley's article "Predicting college success of the educationally disadvantaged" (19 Feb. 1971, p. 640) where he discussed, among many other studies, one by Cleary (*1*). Clark and Plotkin took out of context a single sentence in the 7½-page article: "Cleary tried to replicate the findings of Clark and Plotkin [*2*] with a better controlled design, but failed." This conclusion referred to Stanley's prior quotation from the Clark and Plotkin study: ". . . Clark and Plotkin . . . had reported results of a study based on 'alumni' classes of the National Scholarship Service and Fund for Negro Students in which they concluded that:

> . . . scholastic aptitude test scores are not clearly associated with college grades. It is suggested that college admissions officers weigh test scores less, since they do not predict the college success of Negro students in the same way they do for whites. This study indicates that motivational factors are probably more important than test scores in the demonstrated superiority of Negro students in completing college."

Stanley was not questioning their conclusion that an able, highly motivated group of black students persisted well to graduation in a variety of interracial colleges during the 1950's. (For example, see note 19 in Stanley's article.) He did, however, cite much evidence—including Cleary's study—that Scholastic Aptitude Test scores and high school records tend to predict the college grades of blacks at least as well as they do those of nonblacks. In their letter Clark and Plotkin disregard this other evidence and thereby imply that Stanley's whole case rests on the Cleary study alone, which they seem to perceive as part of an Educational Testing Service plot against them ("Accustomed as we have become for our study to be the launching pad for ETS papers. . ."). Although one need not defend the professional integrity of researchers at ETS (*3*), we note that neither Stanley's article nor most of the reports he cites were done at ETS. Some of the strongest evidence came from investigations by black non-ETS researchers such as S. O. Roberts, Joseph P. McKelpin, and Charles Leo Thomas and from studies in Illinois, Maryland, Georgia, and the Seven Sisters colleges. Even a glance at the bibliography in the Stanley article would indicate the broad data on which he based his conclusions. (Indeed, only three lines of his paper were devoted to the Cleary article.)

Since Clark and Plotkin have chosen to criticize the Cleary article, let us consider the points they make. Clark and Plotkin say that their data were better than Cleary's because they "had data on over 1200 Negro subjects (with questionnaire response from over 500) drawn from all sections of the country and distributed in hundreds of diverse colleges." The quality of data is, of course, always relative to the purpose they are to serve. For the conclusion of Clark and Plotkin that an admissions officer (who necessarily works within a single institution) should weigh test scores less, distribution of the sample among "hundreds of diverse colleges" is a disadvantage rather than an advantage. For the admissions officer, Cleary's within-college analysis is clearly more relevant. Since Cleary was able to analyze data in only three colleges, she limited her conclusions: "The schools used in this study do not represent the full spectrum of colleges in the United States, so general conclusions cannot be reached." But there are many studies other than Cleary's in which the within-college analysis has been performed, and most of these were reviewed in the Stanley article. With each new analysis, the Clark and Plotkin conclusion becomes less credible.

In their criticism of the Cleary article, Clark and Plotkin persist in their emphasis of correlation coefficients to the exclusion of regression lines. It is well known that the size of a correlation coefficient is a function of the variability of the group: when the range of scores is restricted, the correlations are attenuated. For this reason and others, the comparison of regression lines is a more appropriate analysis. Clark and Plotkin computed no correlation coefficients or regression equations, nor did they even work within colleges. Instead, they pooled college grades from 187 different colleges and

universities (*2*, pp. 58–59), thereby treating grades at highly selective institutions as equivalent to grades at unselective ones. (We recognize that they had no choice in this matter; at the time of their study, institutions of higher education were so severely segregated that a useful within-college analysis in integrated colleges was impossible.)

Rather than focus on correlations, Cleary proposed a definition of test bias that was to be examined in her study (*1*, p. 115):

> A test is biased for members of a subgroup of the population if, in the prediction of criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or low for members of the subgroup. With this definition of bias, there may be a connotation of "unfair," particularly if the use of the test produces a prediction that is too low. If the test is used for selection, members of a subgroup may be rejected when they were capable of adequate performance.

She was able to conclude that, in these colleges, the SAT was as appropriate for the prediction of college grades of blacks as of whites (*1*, p. 123):

> In the two eastern schools, there were not significant differences in the regression lines for Negro and white students. In the one college in the southwest, the regression lines for Negro and white students were significantly different: the Negro students' scores were over-predicted by the use of the white or common regression lines. When high school grades or rank-in-class are used in addition to the SAT as predictors, the degree of positive bias for the Negro students increases.

In their letter Clark and Plotkin state that one has to examine Cleary's tables to find the differences in correlations for blacks and whites, because she fails to mention them in her text. In fact, in both the earlier *Research Bulletin* (*4*) which they cite and the journal article (*1*) cited by Stanley, three paragraphs are devoted to correlation coefficients; the differences are pointed out and explanations are proposed.

Clark and Plotkin are correct that there may have been some misidentification of race in the Cleary study: in only one school were records of race available from the college; in the other two colleges identification was made from photographs, with corroboration provided from an NAACP list in one of the colleges. Gross errors in classification would be required, however, to change the results markedly.

*Testing and Fair Employment* (*5*) is cited by Clark and Plotkin to indicate that differential racial validity is a problem in industrial settings. This is really not relevant to the question of the SAT in educational prediction. Even so, several important points may be noted: (i) the book routinely examines correlation coefficients, rather than regression lines; (ii) all but one of the studies has very small samples of Negroes, usually about 31, and yet large numbers of correlation coefficients are computed and compared; (iii) in the only study with a reasonably large sample (98 Negro and 437 white), differential validities were not found; and (iv) in the one educational study that they say contradicts Cleary's results, the Negro and white students were in different schools, although they took the same criterion tests (State Examinations in Nursing).

In two recently published studies of educational predictions, Temp (*6*) and Davis and Temp (*7*) argue strongly for validity studies in all institutions that are using the SAT (or any other admission tests or predictors) for the selection of students. We agree with this recommendation. The College Board provides, through the Educational Testing Service, free validity studies to any institution that uses the SAT in its admissions process. Users of the SAT are encouraged to study the separate regression lines for Negro and white students (as well as other groups such as males and females) so that they can make reasoned judgments about the utility of the test for groups of students in their institution.

T. ANNE CLEARY
*College Entrance Examination Board, 888 Seventh Avenue, New York 10019*
JULIAN C. STANLEY
*Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218*

**References and Notes**

1. T. A. Cleary, *J. Educ. Meas.* **5**, 115 (1968).
2. K. B. Clark and L. Plotkin, *The Negro Student at Integrated Colleges* (National Scholarship Service and Fund for Negro Students, New York, 1963).
3. For a rejoinder to one of Clark and Plotkin's complaints about ETS, see R. O. Fortna, *Science* **175**, 706 (1972).
4. T. A. Cleary, *Research Bulletin RB-66-31* (Educational Testing Service, Princeton, N.J., 1966).
5. J. J. Kirkpatrick, R. B. Ewen, R. S. Barrett, R. A. Katzell, *Testing and Fair Employment* (New York Univ. Press, New York, 1968).
6. G. Temp, *J. Educ. Meas.* **8**, 245 (1972).
7. J. A. Davis and G. Temp, *College Board Rev.* **81**, 4 (1971).