Letters

Aptitude Test Bias

For eight years we have refrained from responding to critics of our study *The Negro Student at Integrated Colleges* (National Scholarship Service and Fund for Negro Students, 1963), because we felt that psychologists and other specialists in educational testing could weigh the admittedly conflicting data on the degree to which Negro performance in college can be predicted from scholastic aptitude tests. One sentence containing three egregious misstatements, in J. C. Stanley's article "Predicting college success of the educationally disadvantaged" (19 Feb.), has compelled us to change this policy. The sentence (p. 641) reads: "Cleary tried to replicate the findings of Clark and Plotkin with a better controlled design but failed." We should like now to set the record straight for those readers of *Science* who have not read the papers contrasted by Stanley, and to counter the myth—propagated for years in researches sponsored, as was Cleary's, by the Educational Testing Service (ETS) —that the Clark-Plotkin study was not well controlled.

The most casual reading of the two papers will make evident that Cleary's study (*1*) was neither a replication of ours nor better controlled than ours. Cleary was solely concerned with the predictive validity of the Scholastic Aptitude Test (SAT), which is marketed by Educational Testing Service. The Clark-Plotkin research was only incidentally concerned with the validity of the SAT. We related the college success of Negro students to a host of precollege variables (test scores, high school average, parental occupation and educational level, family income, geographical area, and others). In addition, by means of questionnaires we obtained much information about the postcollege life of our respondents (graduate training, employment record, social attitudes, community involvement, and so on) and retrospective views of their college experiences (for example, bias encountered, attitudes toward faculty and fellow students, college financing, number of hours of employment, and extracurricular activities).

We 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. Cleary's study was limited to 273 Negroes in three state-supported institutions (none identified) in the East and Southwest. Circumstances dictated the samples in both studies; neither achieved the scientific ideal of random assignment of blacks to random colleges without regard to test scores or high school averages. Our subjects came from the files of an organization which assisted Negroes to gain admission to integrated colleges. Cleary could find only three integrated schools in the United States with a sufficient number of blacks to study (*2*). Cleary had a white sample for comparison; we used normative data for whites.

Both studies are limited because of the sample bias inherent in a concurrent-validity design. In our opinion, however, the sample of colleges in the Clark-Plotkin study is far superior to that in Cleary's. It is more difficult to judge the representativeness of the student samples but, again, we believe ours is more representative than hers on the basis of number, geographic distribution, and the longer time span covered in our study.

In validity studies the criterion index is crucial. Cleary used only students in college and thus was limited to the grade point average; furthermore, most of these averages were based only on the freshman year. In contrast, we obtained transcripts after graduation and had two indices of success, graduation itself and cumulative grade scores for four years for those who were graduated. The reader can judge which method is more reliable and valid.

A final important difference is that whereas we knew all our students were Negro, Cleary established race in two schools by judgments of photographs supplemented by NAACP membership lists. Some light-skinned Negroes not in the NAACP chapter may very well have been placed in her white group.

The great differences in the two studies necessitate great care in any comparison of their findings. Cleary avoided this; instead she criticized our study on the following grounds: (i) the sample was highly selected; (ii) the colleges varied in selectivity; and (iii) the same weight was given to grades from different colleges. Each of these statements is equally true of Cleary's study. In addition, she repeated a minor criticism voiced earlier by another ETS staff member, that it was difficult at times to determine which subsample we used for a particular comparison (*3*).

Let us now consider Stanley's statement that Cleary failed to find what we did. Her school 1 data reveal that while the SAT verbal scores were equally predictive for the two races (.45 and .47), the SAT mathematics score for Negro students correlated a mere .01 with grades whereas the correlation for matched whites was .25. In our study (which did not present either validity coefficients or regression lines), we noted that the SAT verbal scores were more associated with college success than were the mathematics scores.

Similarly in school 2, on *both* parts of the SAT the validity correlations were lower for Negroes than for whites. Cleary neglected to point this out in her text, which simply dismissed her findings in this school because there were no "impressive" validities. Since the problem ostensibly under investigation was differential racial validity, the magnitude of the correlation is irrelevant and one would expect at least a sentence to the effect that the SAT was not predictive for either the white or the Negro students.

Only in school 3 were the two racial groups similar in validity correlations. Again, however, one must refer to Cleary's tables, not her text, to discover that the black students in school 3 (the Southwestern one) were far inferior to those in schools 1 and 2 (the differences averaged 150 points on the verbal section of the SAT and 125 on the mathematics section). The following criticisms of this part of the study can be listed as follows: (i) the sample was "lowly" selected; (ii) the colleges varied in selectivity; and (iii) the same weight was given to grades from different colleges.

Since the two schools where the findings were not inconsistent with ours were in the same SAT range as our sample, Cleary's one empirical finding which challenges us may reflect possible nonlinearity of the relation between validity coefficients and the level of the SAT.

In conclusion, it should also be pointed out that when tests are used for employment selection, the evidence clearly indicates that differential racial validity persists as a problem (*4*). Recent court decisions stemming from the Civil Rights Act of 1964 make it mandatory to compute validity coefficients for both races. This is a practice long avoided by test publishers. Sophisticated models of statistical analysis to handle differential racial validity now appear in theoretical journals (*5*). Our study, which opened up a new phase in the discussion of test bias (*6*), is certainly outdated; nevertheless, the evidence presented in it has not been really challenged by either Stanley or Cleary. With future studies and more comprehensive data to be expected from the Open Admissions Policy at the City University and other programs which have increased the Negro ratio at integrated colleges, it is premature to insist that the SAT is as valid for blacks as for whites.

KENNETH CLARK
LAWRENCE PLOTKIN
Department of Psychology, City College of the City University of New York, New York 10031, and *Metropolitan Applied Research Center, 60 East 86 Street, New York 10028*

References and Notes

1. T. A. Cleary, *Research Bulletin RB-66-31* (Educational Testing Service, 1966). Stanley did not cite this, the original paper; he referred to journal articles by Cleary which were published two years later. Most of the criticisms of the Clark-Plotkin study do not appear in the latter.
2. This situation is changing; although social policy is rarely based on a single study, the Clark-Plotkin finding that black students had a higher rate of graduation than whites despite their lower SAT scores and socioeconomic status contributed to the improvement of Negro representation at integrated colleges.
3. J. Campbell, *Research Bulletin RB-64-34* (Educational Testing Service, 1964). Campbell, in fact, in a theoretical paper with no new data, was the first to criticize the Clark-Plotkin research; Cleary merely repeated him. Neither Cleary nor Campbell ever wrote to us for clarification of the points they found obscure.
4. J. J. Kirkpatrick, R. B. Ewen, R. S. Barrett, R. A. Katzell, *Testing and Fair Employment* (New York Univ. Press, New York, 1968). The one study in an educational setting reported in this valuable work explicitly contradicted the findings reported by Cleary. Selection tests for nursing schools (similar in content to the SAT) were found to be more valid for whites than for Negroes when correlated with criterion scores on state licensing examinations.
5. H. J. Einhorn and A. R. Bass, *Psychol. Bull.* **75**, 261 (1971).
6. Accustomed as we have become for our study to be the launching pad for ETS papers, we were shocked to find that in the latest bibliography of test bias compiled by ETS (*TM Reports No. 2,* 1971) our study is not listed. Our problem now is to decide which is worse, misrepresentation or oblivion.

University Organization

Dael Wolfle convincingly writes of the need to alter, rejuvenate, and expand the universities' relict departmental system (Editorial, 9 July, p. 109). Scientific research and education must go beyond reductionism if they are to help us understand and solve complex problems. We need to form new academic structures and to modify the existing ones to better serve the current needs of scholarship and society.

Although a few new schools are being created with different organizational plans, most of us work, and will continue to work, within an existing framework of traditions, customs, and habits that tend to inhibit innovation. I would like to make some suggestions, prompted by my own experience, for moving toward a generalist approach.

1) Changes may be initiated at any level. There is no need to wait for the organizational structure to be changed before moving in new directions. A seminar may be guided into an examination of cross-disciplinary topics without the slow debate of administrative committees. At the University of Rhode Island, a course entitled "Science and Society" was offered as a "special problems" course after it had been tabled by deans who could find no mechanism for approval outside the instructor's unwilling department.

2) Faculty members should meet with colleagues outside their own departments who express an interest in interdisciplinary themes. Informal discussion groups may lead to collaborative research and teaching. Four faculty members joined to teach an experimental course on "people in cities" at our university and stimulated student and faculty interest. Persons who take part in such activities, however, must be prepared to work harder than usual and often must volunteer extra time. Departmental chairmen are unwilling to give released time for such teaching because it threatens to dissipate the human resources at their disposal.

3) The AAAS could encourage and foster adaptive change by following Wolfle's editorial with assessments of new procedures and organizational forms that are being tried. Communication of such findings could be accomplished at national meetings, by publication of examples in *Science,* and by the establishment of a reference center for university reform. While we wait for administrative superstructure to change, I urge that individuals get on with their efforts. Structural adaptations may follow thoughtful, hard work by scientists and professors who take their social responsibilities seriously. Departments may even be improved from within.

GARRETT C. CLOUGH
Department of Zoology, University of Rhode Island, Kingston 02881

Wolfle lists several reasons why the American university needs to be reorganized. He correctly points to the greater effectiveness in research and teaching of the American departmental structure over that of the German system, which is characterized by authority vested in the individual professor. However, in recent years many American university departments have evolved, as a result of the expansion and specialization of knowledge and dependence on external sources for research funds, into loose assemblages of individual faculty members, each of whom has carefully defined intellectual interests. In effect the result has been the recreation of the German model under the umbrella of the department.

In the Federal Republic of Germany, on the other hand, the limitations of the Humboldt tradition have become increasingly apparent. This has led to recommendations of the Science Council (*1*) and several state legislatures (*2*) for the creation of disciplinary regions (Fachbereiche) within the universities. These disciplinary regions would assume responsibility for the execution of research and teaching within large and flexible frameworks. Furthermore the implementation of the disciplinary regions structure has been encouraged through the initiation of a publicly financed interdisciplinary research program (Sonderforschungsbereiche) managed by the German Research Society.

If this attempt to make the German university more effective and responsive is a success, Germany will once again have provided a model for American graduate education.

BERNARD R. STEIN
6727 Rosewood Street, Annandale, Virginia 22003