DOCUMENT RESUME

ED 080 558 TM 003 063

TITLE	Papers Presented at the Graduate Record Examinations Board Research Seminar at the 12th Annual Meeting of the Council of Graduate Schools.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	May 73
NOTE	15p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement Tests; Aptitude Tests; Conference Reports; *Graduate Study; Minority Groups; *Research Projects; *Seminars; *Standardized Tests; Test Bias; Test Construction; *Testing; Test Interpretation
IDENTIFIERS	*Graduate Record Examinations

ABSTRACT

The three papers provided here were presented at the GRE Board Research Seminar: (1) "Background, Purpose, and Scope of the GRE Board Research Program" by Bryce Crawford, Jr.; (2) "Predicting Success in Graduate Education" by Warren W. Willingham; and (3) "Research on Testing and the Minority Student" by Ronald L. Flaugher. (For related documents, see TM 003 064-065, TM 002 964.) (KM)

PAPERS PRESENTED AT THE GRADUATE RECORD EXAMINATIONS BOARD-RESEARCH SEMINAR AT THE 12TH ANNUAL MEETING OF THE COUNCIL OF GRADUATE SCHOOLS

Educational Testing Service
Princeton, N. J. 08540

May, 1973

TABLE OF CONTENTS

Background, Purpose, and Scope of the GRE Board Research Program

Bryce Crawford, Jr.

University of Minnesota

Paper presented at the GRE Board Research Seminar, at the 12th Annual Meeting of the Council of Graduate Schools, November 29, 1972, New Orleans, Louisiana

It is a great pleasure for me to meet in this delightful city of New Orleans with my sometime colleagues in the graduate dean racket; and moreover it is a real satisfaction for me to report on the development of the research program of the GRE Board, which has been a most fascinating concern of mine for some four years now.

Moreover it is entirely fitting that a report be made to what might be called the constituency of the Graduate Record Examinations Board, and no better occasion could be thought of than a Council of Graduate Schools' meeting. The GRE Board has a very close relation indeed to the community of graduate deans, both in a common interest, and in actual tight control. For the Board is constituted basically by appointments from the AGS and CGS, and this relationship is underscored by the fact that there is formal reporting from the Graduate Record Examinations Board both at the AGS annual meeting and at the CGS annual meeting. In addition to this formal control and communication, the GRE Board feels the need for reporting to its basic constituency and parent body in any convenient and useful way. So we have the GRE Board Newsletter, which goes out bimonthly to over 17,000 people in the graduate community; and there is a continual offering of summary reports and other information on activities and changes and hopefully improvement. If any of you find that somehow you are not receiving the Newsletter and similar information, I hope you will get in touch with the secretary of the GRE Board, Miss Maryann Lear, who will certainly see that you receive these communications.

It could be said that it is the intention of the GRE Board not only to act as the arm of the community of graduate deans which seeks to maintain and develop appropriate help in the process of admission and counseling of new graduate students, but to act in a real sense as the "research arm" of the graduate dean's community on the broadest basis. I might underscore this assertion by recalling that in the last year there was a survey of enrollments of new students in Graduate Schools, sponsored jointly by the CGS and the GRE Board, and carried out by ETS under the terms of its relation with the GRE Board. It is therefore in this context of the GRE Board as in some sense a "research arm" of the graduate

dean's community that I would like to describe to you the scope and range and purpose of the GRE Board Research Program.

It's worth noting how any sort of research program of the GRE Board got started, and this involves recalling how the GRE Board itself got started. In the beginning, which is about a dozen years ago, when God through his mysterious and rather questionable means created the Council of Graduate Schools, there preexisted the Graduate Record Examinations, operated by the Educational Testing Service, a nonprofit organization with headquarters in Princeton. There of course also preexisted a Committee of Testing of the Association of Graduate Schools; and the nascent CGS itself rather quickly set up a similar committee on testing; for some sort of instrument is essential in the pragmatic operation of selection and counseling of graduate students. The GRE was operated by ETS, a group of extraordinarily competent psychometrists living in their ivory tower in Princeton; they had masterful expertise with regard to the construction and operation of aptitude and achievement tests, and they had almost complete isolation from the pragmatic decisions in using such test results to guide the admission and selection and counseling of graduate students. The AGS and CGS committees of graduate deans, on the other hand, had the responsibility for making these pragmatic decisions, for which they needed every bit of help they could get; and by and large they had very little knowledge, though rather deep suspicion, with regard to such tests as the GRE. Nevertheless, because the GRE as then operated was one of the very few objective and nationally normed instruments available, it was in fact rather widely used, and this of course made it important.

It was therefore a wise step to bring forth the GRE Board, and we owe much to the academic statesmen in AGS and CGS and ETS who did this. The Board, appointed through the AGS and CGS, controls the policy and operation of the Graduate Record Examinations. It of course has both fiscal and policy relations with the Educational Testing Service, which is in one sense the owner of the GRE and in another sense the operating arm of the GRE Board; all of these interactions are set forth in a memorandum of agreement which I shall not take the time to spell out, but which is certainly available to any of you who would like to read it and study it.

Now the GRE Board itself began operations in May of 1966; and you will understand there was a good deal of time required to get relationships straight, and quite a bit of time and a certain amount of argument and even confrontation in order to achieve mutual understanding and respect between the deans who had to make the pragmatic decisions, and the psychometrists who really knew what could and could not be done

in terms of achievement and aptitude tests. In the first two or three years of the GRE Board, a great deal was accomplished with regard to improving the mode of operation of the GRE so that it could be more effectively used by graduate deans and by departmental admissions officers in their day-to-day practical decisions. But even at the start the GRE Board recognized that aptitude and achievement tests did not cover all the factors which one would like to take into account in guiding graduate students, and that some sort of research program was necessary to improve the overall operation. And in the birth year 1966 the GRE Board authorized its first research project, setting up and authorizing the expenditure of some $85,000 to investigate the possibility of measuring creativity. Let me say at once that this particular project was a fairly complete turkey, and no great useful results came out of it; yet I think that it was a good thing, in that it recorded in the clearest possible way the appreciation by the nascent GRE Board of the fact that the GRE was an incomplete and imperfect instrument, and that it needed improvement and supplementation. That appreciation of imperfection, and determination to improve, has remained the dominant characteristic of the GRE Board, and of course is responsible for the existence of its research program.

I've indicated that the first couple of years of the GRE Board were fully occupied in a general review and revision and reworking of the existing GRE operation; and while a number of improvements and changes were made, there was no formal research program that was instituted until the latter part of 1968; and the Research Committee of the GRE Board was created at that time, and held its first formal meeting in January 1969. Beginning with the year 1969 we find a continuing series of research projects proposed, developed, authorized and funded, through the careful scrutiny of the Research Committee and appropriate action by the entire GRE Board. First, there was a rather deep consideration of the breadth and depth of research which the GRE Board considered appropriate to its purposes. These considerations resulted in the setting out of a document, affectionately referred to as the "Manning Map," which indicates the range of investigation deemed appropriate by the GRE Board. The document is available, and I commend it to you as a stimulating and thought-provoking document indeed. In it the GRE Board recorded itself as interested not only in the improvement of the GRE, or in aptitude or achievement tests, but rather in the broad field of questions concerned with the identification of students who would benefit from postbaccalaureate education, their attraction to the best schools for their improvement and education, their selection, and their counseling.

In that framework we've developed a research activity which by now I think can honestly claim the status of a coherent and productive research program. Directions of exploration and individual projects are hammered out between the GRE Board, and most particularly the Research Committee

and the psychometrists of ETS, and project progress is monitored in the same way. There are formal agreements on policies, particularly with regard to the ETS as the principal research arm of the GRE Board, with appropriate provision for utilization of non-ETS researchers when this is beneficial; if any of you would like to see the formal statement of this policy it is open to you. With regard to financial support of the research program, the guideline is that the GRE Board will set up each year a research appropriation of 5 or 6% of the previous year's income budget; from this appropriation the Research Committee can fund small projects on its own, but must obtain approval from the entire Board for large-scale projects. Without boring anyone with unnecessary details, it is an indication of the seriousness of the research endeavor to point out that, since the GRE income budget runs on the order of $5 - 6,000,000 per year, the research projects to date have involved the appropriation of just a bit over $1,000,000.

Any research program is really one of R & D, and the line between research and development is sometimes fuzzy both in budget and intellectually. A number of improvements in the GRE have involved research projects along the way. There was an initial rescaling, a technical matter that I think we probably should not go into here. More importantly, there is underway at the present time a most significant restructuring of the GRE which takes advantage of the progress in test construction, to use more effectively the time spent by the student in taking the examination. This restructuring has gone forward over the past two years in particular, and beginning with the next October administration of the GRE, will provide more significant and more detailed information in the form of subscores on about half the advanced tests. Again, I invite you to ask for information with regard to the specific improvements and the specific new types of information which will be available from the several GRE tests, in particular the so-called advanced tests in various fields.

I might say that, here as in all research programs, things don't always come out as nicely as one hopes they will. There was a general desire in the GRE Board, and in the various committees of examiners who preside over advanced tests, to add to the so called verbal-aptitude and quantitative-aptitude score something which would have to do with the ability to think logically. This seemed possible; and research was undertaken to develop such a separate section of the Aptitude Test which would have to do with the ability to think logically. Somewhat to my surprise--although of course I can now offer ex post facto reasonable explanations--this test showed an extraordinarily high correlation with the verbal-aptitude test. Setting out the implications another way, what seemed to be a good test for logical thinking turned out to depend very heavily on verbal aptitude. It may well be simply that, in the mode of thinking

which governs the human mind, verbal aptitude is so central that any logical thinker is dependent on the ability to verbalize, or perhaps we should say conceptualize, his thoughts. At any rate, the proposed "logical thinking" module gave no new information. I'm not ashamed to confess this particular failure, because there are a number of successful improvements in which we can take pride on behalf of the GRE Board operation--and our R & D program is continuing.

We have then some 30 or so different projects which have actually been undertaken and funded by the GRE Board; a few of them have by now been completed, perhaps a third, and some of them have been the basis of reports, which all of you can learn about through the GRE Board Newsletter, and which are available to you for the asking. I'd like to briefly survey the pattern of these projects. Roughly they can be divided into three groups, of comparable order of involvement.

The first group of studies has to do with what I might call product or process or marketing improvement, if you'll forgive a chemist's terms. They deal with technical matters and with the very direct utilization of the GRE. Again, not all are successes: one fairly clear technical study on the possible benefits of changing the option weighting in the GRE brought forth the perfectly sound technical conclusion that this would result in an increase in the reliability of the test results, but a reduction in their validity. I imagine that the number of you who want to hear about these technical results, or who find this particular statistical jargon enlightening or interesting, is small. Yet we can all realize the need to carry out such studies if indeed we are to continually improve the actual technical aspects of the GRE.

But I would also include in this part of our projects certain studies which go beyond any narrow concept of the GRE itself. Thus, starting in 1969, we carried out a field survey of actual admission policies and procedures in a small but representative number of Graduate Schools, involving not only use of the GRE (which indeed was the least part of the study) but the general question of the ways in which admissions officers made the best decisions they could make. This study resulted in some Workshops on admissions procedures which proved to be very useful indeed. Further development of this particular project has lead to the compilation and publication of the Graduate Programs and Admissions Manual, which I think all of you will agree is a remarkably useful compilation, and which also shows promise of further development into an even more useful tool, both for graduate admissions officers and for those involved in the counseling and guidance of undergraduate students as they approach the question of which, if any, Graduate School they should think of.

In the same broad-ranging fashion, the GRE Board has not only carried out, in collaboration with the National Research Council, some studies of the usefulness of the GRE, and possibly modified ways of using it; it has also funded a study now in progress which involves following a group of students as they emerge from undergraduate years and go on into graduate or professional studies. Both of these will be reported on in some depth by the other speakers in this seminar so I shall say no more about them. I would however like to leave you with the point that, even in what we can call the "Nuts and Bolts" part of our research program, the GRE Board has gone considerably beyond the idea of examining the GRE itself, and is actively investigating all ways in which admissions decisions can be improved.

A second third, roughly, of our current research program is one which is addressed to the problem of "social justice," as I call it. It is generally felt, though it has certainly not been scientifically proven, that the GRE--as well as other tests--has an intrinsic cultural bias which makes its use unfair to minorities; and there is also the nonminority group known as "Women" who raise the question of possible bias. Here again, in the view of the GRE Board Research Committee, we are interested not only in the question of possible bias existing in the GRE itself, but in the actual operation of admissions procedures in the graduate schools. Even in 1969, the GRE Board, in collaboration with the CGS, carried out a survey on what was being done by graduate schools with regard to disadvantaged students. Since that time, we have an increasing program of studies having to do with the determination of possible bias in the GRE Board itself, and with the determination of ways to eliminate that bias if possible or to correct for it if it cannot be eliminated. These studies include a whole group of projects; again I will say little about these since they will be the subject of one of the larger reports in this morning's program. This whole section of our research program, in my opinion, constitutes a responsive and responsible attack on one of the major flaws of graduate education at the present time.

The third large component of the overall GRE Board Research Program has to do with what we might call "basic" or at least "long-range" research. I mentioned that even in its first year of life, the GRE Board indicated its belief in long-range improvement by funding a study on "creativity." This was not particularly successful; it was indeed premature. But, beginning in 1970, the GRE Board began to fund some projects which, though they certainly had to do with the technical aspects of the GRE itself, can only be regarded as long-range; for they involve deep lying studies on the applicability of unusual types of statistical approaches. These were begun in 1970; I have not yet seen any final reports; but those final reports, when they come in, will only point the way to developments which cannot give us any actual fruit for some years to come.

A little later in 1970 the GRE Board funded a long-range study with the idea of seeking further information on just what went on in the development of graduate students: just when it became possible to say that a graduate student would clearly succeed--or would clearly be a failure. This particular study hit upon the approach which we now refer to as "the critical incident," and we have now underway a study which I believe may be very significant indeed, and about which I will say no more since it too will be spoken of a little later this morning.

But even further: within the last couple of years the Research Committee felt we had reached the stage where we needed to go beyond all of these aspects to see if we could begin to get some handle on the characteristics of an individual, beyond those susceptible to some type of measurement by existing tests, which profoundly affect his performance in Graduate School or in his further career. We consulted with, and argued with, the staff at the ETS: and out of all this we've begun some of the most far reaching projects in our overall research program. I'm referring to the studies on cognitive style, about which you will hear later on in the morning, and which I believe you also will find very interesting.

All in all, I think that we can characterize the GRE Board Research Program as dealing not only with the constant improvement of the present GRE instrument, and the best use of such instruments, but with the whole matter of what we can do to aid the educational officers of both graduate and undergraduate institutions, as they seek to advise and counsel and guide young men and women--from whatever cultural or ethnic group--in the maximum development of their capabilities. And I'd like to underscore the point that I did not say "intellectual" or "academic" or "scholarly" capabilities. We need to keep our eye on these cognitive facets, but we also need to broaden our area of concern, and our area of effective measurement and evaluation and guidance, far beyond this narrow sector.

Research on Testing and the Minority Student

Ronald L. Flaugher

Educational Testing Service

Paper presented at the GRE Board Research Seminar, at the 12th Annual Meeting of the Council of Graduate Schools, November 29, 1972, New Orleans, Louisiana

Whenever the two topics of minority students and objective testing appear in conjunction, a third topic, that of "bias," soon appears as well. Superficially a straightforward concept, it soon becomes apparent that the term possesses enormous complexity, overladen with an emotionality that greatly reduces the likelihood of progress in untangling that complexity.

Some investigators have made valuable attempts to extract the emotionality by careful definitions, largely of a statistical nature, and these will be referred to later in the paper as we review that aspect of the research that has been completed on the minority student. But I would like to try out a much broader definition of bias, and even give that definition a broad interpretation, and let that serve as the organizing theme of a very quick scanning of the research literature.

Essentially, there seems to be some real value in defining the term bias simply as inaccuracy in measurement. The inaccuracy is of a special kind in this case, in that it is systematic and focused on particular subgroups, ethnic subgroups, of the population taking the test.

Now, giving this broad definition of "bias" a broad interpretation, it is interesting to consider the wide variety of sources from which negative influences can come, making themselves felt in the form of increased error in the measurement of persons who are members of ethnic minority groups. Although the content of the test, that is, the test item itself, is the first thing to come to mind at the thought of "bias," this can be seen to be just one of a number of possible sources of inaccurate measurement. Besides the test content, other potential sources of inaccurate measurement lie in the testing program itself; that is, the practices and policies surrounding the delivery of that test content to the student-candidate. In addition, by stretching the meaning of measurement a bit, we can include the actual utilization of the testing information as a source of inaccuracy, in that over-interpretation, or improper application, of the data can represent just as grievous an error as those from other sources. The organizing theme, then, for the research review which follows, is inaccuracy from the sources of content, program, and utilization.

The content of the test can be a source of inaccuracy either because of what it includes, such as questions for which some students have not had an opportunity to prepare, or because of what it does not include, such as those topics or strong points possessed by the student other than perhaps the traditional verbal and mathematics facility. Theoretically, this question can be settled easily by simply referring to the predictive validity of the content as the determinant for its inclusion in the test. If the student has not been exposed to the content, this is what we want to know; this is what the test is supposed to be finding out. The argument could be similar for that content which is not included: if it doesn't relate to the criterion of school performance, don't include it.

But the reality of the matter is not so easily handled; minority spokesmen could claim that there are a large number of possible items, equally valid as predictors, but with differential difficulties for minority versus majority students; and as for new content, there are many different kinds of aptitude tests that have never been studied as valid predictors for minorities.

The study of differential item difficulty lacks a really satisfactory method for determining just what constitutes an "unreasonably difficult" item--although there have been several elaborate attempts--and leave untouched the question of how the predictive validity is affected. Meanwhile, test constructors are, in fact, including particular items that demonstrate the awareness of the interests and activities of minority groups; but the research findings have been of little help in guiding these changes. Similarly, for the possibility of adding other kinds of measures to the traditional verbal and math scores, the research data are just not available.

The reason for this inadequate state of affairs might well be the lack of success of the several initial explorations into these questions, whose findings have been rather universally that, under the most extreme circumstances detectable, the totality of the indicated changes simply would not make that much difference in the scores of individuals. So the results of our hunt for a source of bias in the specific content of the test have been discouraging.

The second potential source of inaccuracy, that of the testing program itself, can be divided for convenience into two sub-categories, that of "atmosphere" and that of "presentation." Presentation, having to do with the characteristics of the test, other than the content, such as the speededness, or the tests' coachability, has been the subject of several research studies; actually, one is underway currently,

sponsored by the Graduate Record Examinations Board. The other subcategory, called atmosphere factors, however, includes such characteristics as the recruitment policies of the testing program and characteristics of the testing room. Up to now, these have not been very popular topics for research, but hopefully this will begin to change.

It happens that the GRE Board is also sponsoring a study that is one of the rare ones that are investigating the effects of recruiting policies. The evaluation of the GRE Fee-Waiver Program is attempting to determine the success of the attempt to e[illegible] number of qualified but financially disadvantaged students wh[illegible] s[illegible]er graduate school. Research studies such as this one are ways in which the recruitment aspect of the testing program can be checked; there must be constant assurance that groups of qualified minority candidates are not being passed over because they are being discouraged, perhaps inadvertently, through some characteristic of the atmosphere of the program, from attempting to compete as candidates. In addition, a continuing monitoring of the program's descriptive statistics is another good way to assure that the program is attracting the sort of students it wants.

As an aside, while we are on the subject of program descriptive statistics, let it be noted that these statistics are reflections of the success of the recruitment activities, and not somehow norms, representative of the entire ethnic subgroup. Especially in view of such variations as the fee-waiver programs, it should be apparent that the representativeness of that sample is very much in doubt, and interpretation in this manner, though tempting, is dangerous. Descriptive statistics are much more accurately considered as an index of the success of the recruitment program.

As for research on the characteristics of the testing room itself, some studies a few years ago did demonstrate that test scores of minority students do change as a function of such variables; these are difficult studies to conduct, and more are needed, but for practical purposes it is appropriate to assume, even without the hard evidence, that testing room conditions need careful attention if greater accuracy is to be achieved in assessing minority students.

In this same category, the GRE-sponsored study is looking at the question of coachability of the test, or as it is sometimes called, "susceptibility to short-term instruction." If the nature of the presentation of the test material is such that some students are at an advantage for having seen it before, then some inaccuracy in measurement can occur. Coachability is a frequent concern of minority spokesmen, who

argue that part of the reason for lower scores is a lack of familiarity with the "tricks" of taking a test, therefore, the argument goes, it is only reasonable that minority students be given a short course in those things which white middle-class students know from similar training and extensive past experience.

Certainly if such training is possible, it should be provided to everyone, or the nature of the test should be changed so that such training is not productive for anyone. The assumption has always been that all students are equally versed on the medium through which the measurement of achievement or aptitude is taking place. If, in fact, there is some noise in the system, perhaps as a result of becoming confused by the instructions, or some similar non-content factor, unfairness and inaccuracy will result. Research studies that have attempted to cause these large score changes in short periods of time have ranged across a whole spectrum of approaches and levels of care and intensity, with largely negative results. Recently, a successful attempt was reported that was directed toward mathematics items, but the length and intensity of instruction was so great that it almost could qualify as legitimate curriculum material in itself. In addition, the question of relative improvement in scores by minority versus majority-groups was not encompassed by this previous study. The study sponsored by the GRE Board will, however, and should provide important information on this question (Pike and Evans, 1972).

Other aspects of the presentation of the test, such as that of the speededness, have received by now a fairly adequate amount of attention; although each particular test deserves a check on the effects of its speededness, the variety of studies now available do permit a tentative general conclusion about this potential source of inaccuracy. Speededness does not appear to be a major cause of inaccurate measurement differentially for minority versus non-minority students. The score improvements that have been caused by reducing the speededness of a test have been about the same for both minority and majority groups, suggesting that this is not going to be a productive area in our search for inaccuracies.

We must not overlook the possibility that small increments of error from each of several of the factors mentioned above can actually summate to a significant amount of inaccuracy in the test scores of minorities; for example, small inaccuracies from test content might combine with small inaccuracies from speededness; before such interactive effects can be studied, however, we need to complete the documentation of the single sources themselves, such as clarity of instructions, of which we know little.

The third source of inaccuracy, actual utilization of the information that is generated by the testing program, is distinctive because this activity takes place out in the using institutions, apart from, and therefore under only minimal control of, the testing program itself. An amount of responsibility obviously rests with the sponsors and producers of the tests, however, to assure that misuse is at a minimum.

One of the primary questions concerning the use of the test is that of the predictive validity: do high scorers on the test perform better in the curriculum? Comn only heard accusations from minority spokesmen are that the tests may serve well for majority students, but for minorities they are "not valid." If the validity being referred to here is that of a predictive sort, then the appropriate steps for checking on this accusation are obvious; simply that of conducting a study of the differential predictive validity for majority and minority students. Research of this sort has been done across many tests and many curricula, and the conclusion is clear--psychometric predictive validity is about the same for both minority and majority groups.

Difficulty occurs because whenever a term has both a technical and a common usage, and "validity" is such a term there is a potential for confusion. When minority spokesmen proclaim with absolute certainty that a test is "not valid," they may very possibly be using the term in a way that does not correspond completely to the technical use of the term; therefore, when a psychometrically precise validity study is conducted, from which the conclusion is made that the tests are, in fact, "valid" for the minorities in question, such a response may not be an appropriate one to answer the accusation. Lack of validity by common usage may be "proven" by the identification of <u>one</u> person who was turned away, or advised not to attend on the basis of the test scores, but who somehow circumvented the situation and went on, and eventually succeeded. Any procedure that turns away a potentially successful student must be "invalid" in these terms. But, this kind of "proof" of the invalidity of the test, of course, is quite compatible with a simultaneous demonstration of adequate predictive validity by psychometric standards. Even the best of predictive measures necessarily has its share of cases which are falsely predicted to be negative, and this occurs in all ethnic groups; for that matter, there are always cases which are falsely predicted to be positive. Such is the state of art of academic prediction, and this fact may well be the source of confusion and misunderstanding between minorities and admissions officials. In a sense, the difficulty lies in an <u>inflated</u> impression of the effectiveness of testing on the part of such spokesmen, rather than in a belief that they are truly valueless. Any deviation from perfection is proof of invalidity.

On the psychometric side of the dialogue, however, things are certainly far from settled. A few years ago we thought we knew how to determine, with great precision, the fairness or unfairness of a particular test, if only given the proper data. In 1971, Robert Thorndike destroyed our complacency by showing that the traditional study of regression lines was not taking into account an alternative and equally reasonable definition of fairness. He showed that our traditional conception could be fair for a given individual, and yet unfair in terms of the relative proportions of potentially successful students who were selected from the subgroups of the population.

The implications of this dramatic development are still being worked out, and it is too soon to know precisely where additional study will lead, but one possibility would seem to be quite beneficial, and that is the introduction, of necessity, of some non-technical source of the critical decisions. Value judgments implicit in the admission process will have to be made explicit and compared; for example, how much more desirable is it to reject some students who would have succeeded, in the interest of a high success rate in the curriculum, versus the admitting of some students that are likely to fail, in order to ensure that the few successes in that same score range will be given a chance? Are these relative desirabilities different for minority and majority students? Statisticians cannot make those decisions, but they can assist in causing them to operate in the admissions process. Research results to date seem to indicate that this is the appropriate direction for the future.

In summary, I have attempted to point out that there are many other potential sources of bias besides that of the particular item content within the test. The other potential sources, which I designated as program and utilization must also be encompassed in any thorough and effective program to increase the accuracy of assessment for members of ethnic minorities. As usual, the research findings are emerging much more slowly than we would like them to, but that is the nature of careful research. Meanwhile, our failure to find bias from those sources that are most often identified, such as test content, or predictive validities, must not be used to justify an abandonment of the search. The research efforts must encompass these other possible sources, of inaccuracy, too; for that matter, they should be continuous, serving a monitoring function of these possibilities. But meanwhile, we can be aware of them, and of the things that can be done to increase measurement accuracy, based on research evidence or lacking that, just good judgment and sensitivity.