

PANEL: Testing, the Search for Ability, and Equal Employment Opportunity

Preston Munter, M.D., Moderator
William H. Enneis
Everett Belvin Williams
William M. Boyd II

MUNTER: My job this afternoon is to moderate a panel on Testing, the Search for Ability and Equal Employment Opportunity. I can assure you that I will be the fairest possible moderator who could have been chosen for this job because I have sublime ignorance about the processes which go into these things. My only qualification I suppose for doing this job is that as a psychiatrist working in a student health service, I am likely to see both the joys and the sorrows that come as a result of testing and the search for ability.

This is a conference about affirmative action, and the word action interests me always; it certainly interests me when we talk about equal opportunity of any kind because it seems to me that equal opportunity doesn't have any meaning at all unless there is some action. We are all familiar with the wide repertoire of rationalizations that stem from concerns about ability, skills, background, education, training, preparation, etc., as qualifying factors for job hires or admission to schools, for meeting the competition (an unfortunate phrase it seems to me), so any discussion about testing and the search for ability in the equal employment and equal admission opportunities that that suggests must center around those issues and many other issues that are both implicit and explicit insofar as any affirmative action program is concerned.

I would like to take just 30 seconds more to say that the fact that this conference is going on and that we are in this session talking about these matters suggests to me something quite affirmative, and something quite positive that I think is worth pausing over for just a moment. We are in a tough fight—this affirmative action business. It is tougher than it ought to be. I guess I share the impatience of many people in this room about the speed or lack of speed with which changes are being made in this area. On the other hand, I have a very strong affirmative feeling about how much change has already occurred. The fact that we are not talking about whether to, but how to, and discussing it sometimes in a heated way, but very specific ways about what to do to make affirmative action work as a process—it's not an end point, it's a process, it's part of a total social revolution in this country, and the fact that we are talking about ways to make that work, and work better, seems to me to be terribly important and very positive. I don't share some of the negative tone that I have heard so far in the conference, because I really do believe, having been in this a very long time myself—that though we haven't arrived where we would like to be, we have come a very, very long way.

Our first speaker is William H. Enneis. He is chief of the Research Studies Division of the Equal Employment Opportunity Commission. He is

also a professor of psychology at the University of Georgia. He has been a psychologist in various capacities for the United States government, and was a major contributor to a government publication called, Personnel Testing and Equal Employment Opportunity. Mr. Enneis.

ENNEIS: Thank you very much, Dr. Munter. I noticed as I picked up the program here and the list of participants that it looks like Harvard has everybody else out-numbered. I don't know whether there is any significance to that or not, but I must tell you that I didn't come here really to talk completely about personnel selection in the academic area. I am going to, this afternoon, address myself really to what I consider to be the general problem of employers. I don't care who they are, whether they are private or public or whether they happen to be academic institutions of higher learning, or whether they happen to be elementary and secondary schools. I will consider the problem of what you do, to find out whether you have a problem or what you do generally after you find out you have the problem. If you don't have the problem, which I shall define in just one minute, there really is not fundamentally the issue of testing and personnel selection, at least from an equal employment opportunity point of view. There may be a problem in terms of the quality of the people that are being selected for their respective positions. But that's the institution's worry in the fundamental sense, it isn't a legal problem in terms of civil rights.

In fact, one rather leading psychologist in this area Professor Robert Guion has often said that employers are indeed entitled to be fairly stupid in personnel selection as long as they are stupid fairly. What I want to do first of all is to tell you that in an equal employment opportunity context, the first thing that any employer should do is look at the composition of the present work force. After all, that is a known quantity, it is there and it can be studied. And no employer can really honestly say that they don't want to find out what all of their employees are, that they don't care what they are, on some pretense that they select in some equitable way or as the governor of New Hampshire, who defied the Equal Employment Opportunity Commission in filing a report for the entire state. He said he would not file a report to the EEOC because "all the employees of the state of New Hampshire are American citizens and that's good enough for me." Really no employer can make the kind of excuse that they are doing something that is loyal or that they don't intend to discriminate. Every employer can look at the work force and see what it is, see what its composition is with respect to sex distribution in various kinds of jobs and with respect to minorities. And I think you know that Title VII of the Civil Rights Act of 1964 covers classes of persons on the basis of race, color, religion, sex and national origin. The matter of color in fact, in the future may become a very interesting one because there are some people now who say that they are being discriminated against on the basis of their color, for example, the East Indians, who are not Black, but who are in fact as dark as many individuals who are called Black, are regarded as Black in this country. So in fact that may be an interesting point. In fact we have had such charges filed with the Commission, but I am not going into that.

Certainly an employer can look at occupational categories and see how people are distributed in them with respect to sex and various racial or

ethnic categories which in fact are included on the reporting forms which Equal Employment Opportunity Commission has. And as of next fall we will have six of these in operation. The last one, the EEO-6, will in fact be directed to institutions of higher learning and will replace the reporting which they have previously done on the EEO-1 form. I might say that state institutions of higher learning will in fact also be under EEO-6 rather than EEO-4, which is the current public employer reporting system.

Certainly an employer should look for gross under- and even over-representation of groups in various occupational categories. If one finds that 95% of all the employees in a particular occupational category are women, or are Black, I think that over-representation in itself, may be an indication of under-representation in other areas. Many people often forget that over-representation in one job category is a very, very likely signal that there is discrimination and under-representation in other job categories. I might say that the employer should then of course compare the composition of the work force with the relevant labor market. It may be that the relevant labor market contains very few of certain groups and in that kind of situation the employer may be justified, I said may be justified, in arguing of course that the individuals are simply not available for employment. Now I am not going to get into all of the difficulties and arguments that can arise there. That is, the comparison of present labor force with the relevant labor market. But, I think certainly this should be done.

But in addition to proportional representation certainly, I think employers have an obligation to look at the wage and salary structures for individuals and see whether there are large discrepancies between men and women doing the same kind of work, or between Blacks and whites, or between let's say Spanish sur-named Americans and Anglos or Blacks as far as I am concerned, doing essentially the same kind of work. When I say essentially the same kind of work, I mean fundamentally, essentially the same kind of work. I don't mean the fact that they are in different departments because we have found many times that individuals are in fact sex segregated, by job classification, they are segregated on the basis of race or national origin and fundamentally when you look at what they are really doing, even though their job titles are different, and even though they are getting paid quite different amounts of money, they are fundamentally doing the same kind of thing for the employer.

Now, another thing that I think an employer can do, if they find some of these gross discrepancies that might exist with respect to sex, national origin or race, is to then look at the proportions of applicants among these groups that they have for various types of jobs. Fundamentally, I think you realize this is a recruitment problem. You know that many people say, "Well you know those people simply don't apply", or "they don't want to work for us". Now I think that you should probably realize that that sort of an argument really does not cut any mustard today. In fact, discriminatory recruiting systems are a very important part of the whole picture. Those that operate in such a way as to exclude individuals by lack of information being available to those groups of people, or certain groups of people, can be discriminatory. For example, an employer is not off the hook if the employer is relying on word of mouth recruiting by present employees, or

has members of the faculty write letters to their friends, saying hey is somebody good graduating from your department this year and wouldn't you recommend that individual to me, etc. The perpetuation of exclusion can occur in that kind of situation.

In the private sector many employers do say that they like to hire people referred by their present employees because they make good workers, they are loyal and they have a sense of community. I have heard college professors say the same thing, in fact I heard it for four years while I was teaching at the University of Georgia—that it was desirable for the present faculty to try to do as much recruiting as it could because if they did this they would perpetuate the—and don't laugh—the fine standards that the department of psychology had at the University of Georgia. I think you know what is going to happen in a situation like that because obviously at the University of Georgia, and I say obviously, I hope you do know it, when I taught there in the late 1950s, the faculty was of course, exclusively white. There was I think one Oriental professor who was hired by the physics department and was virtually driven out by harrassment.

Also, employers can't rely on walk-ins. I don't think this is particularly relevant for academic institutions at least in the academic part of the employment, but it certainly is for the non-academic positions. Because walk-ins are often dependent on the neighborhood in which the employer is located. I also might say that there are some problems too in referrals from employment agencies when those employment agencies believe that the employer wants a certain type of individual even though the employer has never specified that as such either in writing or in any explicit oral directions.

Well, if one finds that there is a big disparity in terms of the applicants that apply, the employer has got to do something then obviously about the recruiting system to broaden its coverage in terms of sex balance, and also in terms of racial and national origin balance. But going beyond that, another thing that the employer should do, and this is really the third step now that I am talking about in doing an audit—that is an audit that any employer can do—is to do an analysis of the net effect of personnel selection procedures, including promotions and job transfers. First of all the employer should look at the overall rejection rates for every racial, ethnic and sex group, at every stage of the employment process. And I want to say this, that many people now are talking about the end-result theory. They say well you know if we end up with the proper proportion of women, or if we end up with the proper proportion of Blacks or Chicanos, this is fine, because we have met a particular kind of goal. But I find that there is something a little disturbing too in that kind of theory, now called the end-result theory. Because actually that kind of action and this is my own personal opinion, this is not an Agency position, is really very, very dangerously close to the quota system, which of course is expressly prohibited by Title VII. I don't think that an employer is going to really achieve equal employment opportunity on the basis of sex or race or national origin or religion if they hire certain numbers of people and say I have got my numbers, because people are not being hired on the basis of their ability to do the job and I think that is what Title VII is all about. Now

I think there are some other things that Title VII is all about as well, but this so-called end-result theory, if I hire a certain number that's fine and don't worry about anything else. I think this is very, very bad.

For example, I know that one police department said, well we are going to hire a certain number of Chicanos, we have a height requirement though of 5 feet 9 inches. Well it's possible to go out in the Southwest and find Chicanos that are over 5 feet 9 inches, but the proportion of Chicanos that are over 5 feet 9 is much less than the proportion of let's say, Anglos or Blacks that are 5 feet 9 and above. So really that police department was able to hire a certain proportional percentage of Chicanos, because they selectively recruited among them, but what they were really doing was precluding a greater proportion of Chicanos in the community from becoming police officers than Anglos with no evidence whatsoever of the job relatedness of that height requirement. And I say that is why the end-result theory can be discriminatory and that really disturbs me. I have to say something about this because it is something you are probably going to hear about. Employers are going to say if we hire enough people isn't that great. You know. I say no, that's my personal opinion. Because I think it can have a discriminatory result.

Of course, I think perhaps you know that our regulations, as embodied in the *Guidelines on Employee Selection Procedures*, state that if there is an adverse impact on any group covered by Title VII, that then the job relatedness, that is the validity of that selection procedure or procedures must be shown. These steps are outlined in these *Guidelines on Employee Selection Procedures* which are available for no cost from the EEOC if you write them.

I am not going to say much more about that right now, except to also say that the Office of Federal Contract Compliance, who some of you may be responsible to, also has a corresponding order on employee selection on testing and other employee selection procedures. Their order is about 95% identical to ours and in fact, the intent is to have these documents impose the same requirements on those persons who are covered by the OFCC Regulations, that is Executive Order 11246 and 11375, as they may have been amended. In any event, the OFCC does have a corresponding document and the OFCC also has published recently a booklet called "Questions and Answers on the OFCC Testing and Selection Order" that is available from the U.S. Department of Labor, specifically the Office of Federal Contract Compliance, and I commend it to you because I think it is a very good document. It answers some basic questions about employee selection and the procedures that are required for showing the validity of selection procedures.

What an employer does after he finds adverse impact is basically to find out then what people really do on the job. Many employers don't know. That is why in fact some of them have such lousy selection procedures. Some employers I think fortuitously have good selection procedures because they arrived at them by accident. But many of them do have very poor selection procedures simply because they don't know what people do on the job. They select the tests first, then they try to correlate or relate the tests to job performance, and they are surprised sometimes when the tests

do not show any job relatedness, in other words there is no validity. This is not surprising because the selection procedures that they are using are not really related to what people are required to do on the job. The fundamental thing to do certainly in finding out what people are supposed to be doing on the job is to do a job analysis which tells in some detail what those individuals should be doing.

Job analysis is not an arm chair approach to finding out what really the functions are that people perform. One doesn't sit in one's office and just say I have seen somebody doing something over here, and I think this is what is required. I would certainly say don't rely on your personal judgment. I wouldn't even ask supervisors what characteristics they want in their employees because those supervisors may want, or say they need, some very superficial kinds of things that are really not fundamentally related to job performance. I would also warn you not to use the characteristics of the present work force as an indicator of the characteristics that are needed to do the work because many employers will look at their present employees and say my present employees are like this and this is what I want to perpetuate this good work performance. It may be completely an accident that the characteristics of the present work force really are related to any kind of job performance at all. That is a very, very serious mistake. There is a process called worker analysis where you study the characteristics of workers, but that should never be confused with job analysis which is really a study of what people are required to do or what they should be doing. In many cases a job analysis will indicate that people are not doing some of those things that they are supposed to be doing in the job, or could have been doing anyway.

I think that after that is done, the job analysis then should be used, can be used, for many, many purposes. I think the point this afternoon is that we are talking about selection, and I would recommend to you that if you are going to validate your selection procedures that you get somebody who knows what he or she is talking about. Don't get somebody who says, I will do this for you, and don't worry about my credentials. I really think you should use an expert here. I think it is very, very important.

I am not going into the details of test validation this afternoon because we don't have time, but let me close my talk with this point. After selection procedures are in fact validated, and some of them have been shown to be job related, if you have any control over the administration of personnel selection procedures, do not let anybody else pervert the use of those findings. This is one of the biggest mistakes that I have encountered at the EEOC, aside from not doing any validity studies at all. That is, that the users of the test actually pervert the whole process. This issue, for example, arose in a case called *U.S. vs. Georgia Power Company*.¹ First the test being used didn't have very much validity if any, at all, but even so, even in one situation where the validity study seemed to indicate that the selection procedure was *inversely* related to the performance on the job. *Inversely*—that is the higher scorers were doing worse, and the lower scorers were doing better—the company persisted in selecting the people with the higher

1. 474 F.2d 906 (5th Cir. 1973).

scores. In other words they didn't even take the advice of their own consultant and their own expert. Because the man who did the study was certainly qualified for doing that kind of work. But no, they didn't use the results that way at all. The company just turned right around and kept on using the test the way it had before.

I must tell you something about another Equal Employment Opportunity Commission publication which has recently been published. It is called *Job Discrimination? Laws and Rules You Should Know*. I would commend this to you too, because it has all of the guidelines and regulations and the standards of the EEOC in it. I think it is a very valuable book and I would recommend it to you very highly. Thank you very much.

MUNTER: Thank you, Mr. Enneis. Our second speaker is Mr. Everett B. Williams, who is Vice President of Educational Testing Service. He is a psychologist and formerly was Associate Dean for Admissions at Teachers College of Columbia University. Mr. Williams.

WILLIAMS: Thank you. I thought that I would try to just outline a few topic areas in this area of testing with which I am familiar and enter some caveats and some disclaimers here and there and then sit down and have you raise any questions you like. But there are six areas I would like to touch upon. First, there are some definitional matters—definitions of tests and so on. Second, looking at some of the conditions of use. Third is the issue of cultural fairness and cultural boundness. Fourth, some inherent problems in the use functions that are described. Then I would like to briefly try to reconsider at least one of those problems. Then I would like to touch briefly on what would appear to me to be some conflicts between some legal proceedings and maybe some psychometric ones. At least to the extent that people are not talking about the same things.

Just in beginning, for some of us who worked on the OFCC guidelines which, as Bill mentioned, are pretty much the same as the EEOC guidelines, I was always struck by the difficulty in just coming up with some workable definition as to what groups were to be included and those guidelines finally—race, religion, national origin and sex—were settled upon. There is a new category that is under great discussion these days—it is difficult to test for, but it's referred to as sexual orientation, which presumably would take in sex preferences, At least the American Psychological Association has been worrying about that as well as some recent announcements of the American Psychiatric Association regarding the same.

My comments here are geared towards those situations in which I suppose in a university context one talks about non-exempt staff. I mean by that non-faculty personnel since in most universities, and certainly Harvard is not an exception, the peer review procedure is the one that is relied upon most heavily. Also, I want to explicitly exclude the problem of academic selection which was involved in the *Defunis* Case which was talked about earlier in the day. Although we are involved in academic selection—we, meaning ETS—I for the moment just want to concern myself with the employment problem.

The definition of a test that I just want to call to our attention for

discussion purposes, is that we are concerned with getting some sample of behavior which is elicited by some agreed upon procedure and stimuli. The procedures have to be articulated, that is, made explicit, hopefully followed in some sequence, and on the basis of that procedure, those stimuli and so on, there is the intent to draw an inference about some set of behaviors in some other domain of time or location or performance. That means it would cover what most people think about as paper and pencil tests, behavioral tests, performance tests. I would cover things such as medical tests where someone wants to look at or get some reading on blood pressure or whatever else—you do something to get some sample of behavior, some reading on that behavior and draw an inference about some other set of behaviors, either the condition of the heart, the condition of the health of the patient or whatever. That, in most situations, is what is meant by a test. However, in a lot of the discussions in a lot of litigation, there is only one particular form of test that has come under scrutiny, and by and large, those are paper and pencil oriented tests. I think they are because they are very cheap and inexpensive. You can't really get anything more cheap than a piece of paper and a pencil to pick up certain kinds of responses.

There are some other characteristics that are commonly talked about in testing, namely, the validity of the test. By that most of the time people mean the extent to which a test measures what it is supposed to measure. However, not all the time. There is another construct that is used such as reliability, which most of the time means the consistency with which a test measures what it is supposed to measure. Again, not all the time. Other constructs that have come in, such as bias, and bias is sort of a two-edged word in psychometric literature in context. All it means is some systematic response direction, which means that if one were to test for example, as it has been in the past, a group of females with a test that was norm for males, there would be a systematic bias, namely, females tend to score higher on certain paper and pencil tests that sometimes are administered to males. That would be a bias. Now—go the other way, if they scored lower, that would be a systematic bias. But it has no implication of value which I think the social term of bias does.

Fairness is the next term that is used and it is really different from bias. It is somewhat similar to what is used in legal circles, at least in EEOC guidelines and in the law, and has some kind of disparate impact but I will have to explicate that a little bit more. But fairness is used differently from the term bias, at least in testing context.

What are the conditions in which a lot of these things are used. The major function,—at least the one that is most well known, is in the selection function. By analogy admissions to any school, law school, graduate school, undergraduate school, etc., is a selection function. The second major area or function is that of a diagnostic or prescriptive function in which one is interested in knowing how much mathematics does this person know. Or how well does he read? At what level does he read? How rapidly does he read? And once that has been determined, it can then be decided on what course of action in terms of remedial training or some advanced placement might be indicated on the basis of those responses. In the selection function it is usually assumed that there are more people available for something,

whether it is for admission to college, admission to a job, admission to the Army, than there are places. Accordingly, there is a high assumption in our society that when possible you try to make the decision on some rational ground as opposed to irrational grounds. The irrational alternative is sometimes very attractive. For example, suppose you did have more people applying to let's say, Harvard Business School, than there were seats. It is conceivable that you could use a random-number generator and assign them by chance, as long as you filled up the seats. You certainly would have a fair procedure. It would not be in accord with what we usually think of as a rational procedure. But it might, in terms of end-results, yield about the same. That is debatable.

Anyway, the selection function implies that you have to somehow winnow deal with the smaller number of people than there are available. The diagnostic one where you are really trying to understand some process whether in a learning situation such as school, such that you can be more effective in presenting some teaching prescription or material so that some kid or student can maximize his learning, is clearly the more difficult and complex one for a variety of reasons. But it isn't a kind of goal that is attended to as much as the selection functions. One of the impacts I suspect of equal opportunity thrusts will be to give a greater focus on these diagnostic and prescriptive functions.

There are some inherent problems in testing that as far as the literature would suggest there probably is no way around. Every test is culturally bound. You have to present it in some form either in words, or objects, or something. And those objects are usually more familiar to some culture than they are to some other culture. The propositions referred to in sentences in tests usually are more familiar to some sub-set than they are to some other. As a matter of fact, there is one position that would argue that any test that is completely cultural neutral is probably useless. Accordingly, what you want is to try to get some accurate reading on the particular sub-culture with which you are concerned for whatever purposes and objectives that you have in mind.

The problem of tests or cultural boundness is sort of aggravated by the fact that none of the instruments that we now have are free of errors of measurement. There exists no instrument that I know, and certainly in education and psychology and probably in no other area that does not have the problems of errors of measurement. Accordingly, as long as there exists some errors of measurement, and as long as there are sub-cultural differences, it is almost always the case, if you use certain simple predictive systems, that one group that scores lower on the test than another will in some sense be unfairly treated. This raises a kind of interesting dilemma I think, both for psychometricians and maybe for lawyers and those who are implementing EEOC guidelines. That is something that was referred to this morning, as the use of class information—that is whether a person is a member of some class, some identified group, whether it's Blacks, women, whether it's all Chicanos who are 5 feet 9 and over and so on. As long as that kind of demographic information poses a problem, I suspect that the inherent problem which is already a part of testing, is not going to get any better. It is likely to get a bit worse. The reason for that is that in most

testing theory there has always been the presumption that you use all information possible. You try to combine it in some fashion which it will then yield, or permit you to draw, some valid kinds of inference. If you cannot use this information, all it means is that the error factor increases. Error factor in the testing sense means all the things that you don't know that are probably operating to affect a person's behavior, and they are numerous.

Most of the debate about the use of tests, and certainly the intense social feelings about the use of tests, relates to the selection function in which there is the sense that somebody is acting as a gate-keeper, some people get in, some people don't get in. Opportunities are afforded to some and denied to others.

There is another way of perhaps looking at the problem and I would like to perhaps just briefly suggest that. Since the employer and the school or the institution usually are the ones who administer the test, one might consider some paradym that have been used elsewhere, such as Philadelphia Police Department and the State of California. The University of California system has something comparable not quite like this but it is comparable I think in substance, namely, let the individual be both the administer and the interpretator of the test. This would be an alteration where it is possible for you, hopefully, with careful job analyses of the sort that Dr. Enneis was suggesting, to say that this job requires these sorts of things. This is the kind of language that we use. These are some of the meanings. Take this material that I will give you, go home, look it up. Study it. Try to understand the process. When you think you are ready, come back to us and see if you would like to have this job or test yourself and see how well you do. This is not such a strange notion in various professional groups. Actually, there is something comparable that now functions or is in effect with dentists, where they are able to after a period of time being out of school, to self-administer a test and check to see how well they are doing on the knowledge of their professional fields. I suspect this would be helpful for most professional areas, certainly psychology and probably others. If one were to do that, and this has been tried with some success in Philadelphia, one of the things that you pick up that is a little different from the kind of information that you get on the usual testing programs, is some indication of the motivation and interest of the individual over a period of time. He or she is in sole control of the time that they want to spend on the material, studying it. They can bring themselves to a certain state of mastery. And they can decide over a longer period of time is this the kind of thing I want to do. If they decide yes, and you still have more people than you can deal with and they are all qualified, according to this kind of criteria, then it doesn't seem unreasonable to me that you might want to use a random generator of some sort.

In addition, I think it helps individuals to understand more about himself or herself in terms of what they think their vocational interests are. It certainly permits them to feel that they are in control of their own destinies.

I want to just leave that as sort of an unelaborated kind of thought because it poses a real possibility in the employment area where some of the

validity strategies are not as straightforward as they are in some of the academic selection areas.

Let me then just turn to the last area in which I want to just mention are some potential conflicts. I referred to one and that is the use of class data or demographic data in regression formula or any kind of prediction system, whether it is a classical linear regression formula or some other type it doesn't really matter. Since most psychometric theory assumes that you try to get everything that you have, as long as it is contributing to the understanding of the criterion variable.

The second thing that I think is still somewhat troublesome, and some of the speakers this morning referred to it, are the particular validity strategies that are identified in EEOC guidelines, in American Psychological Standards on testing and assessment and so on. The three that were mentioned this morning are those of criterion-related validity, content validity and construct. Interestingly enough, I think that the one that is more favored by some of the government agencies, namely predictive validity or criterion-related validity, is seen by professional psychologists as only a sub-part of a larger one, namely construct validity, and probably many psychometricians would argue that the fundamental approach to the validity question ought to be construct validity and not predictive validity and maybe not even content validity. But this is a kind of debate that swirls around a lot of meetings and certainly the Equal Employment Opportunity Coordinating Council at the moment is having its troubles with it. I would submit that they are not going to be solved primarily in the legal arena because psychologists feel quite strongly about some of their terms in the same way that lawyers feel quite strongly about some of theirs. And I am sure that the notion of construct validity will have to be revisited.

MUNTER: Thank you Mr. Williams. Our last speaker is Mr. William M. Boyd, II. He is president-elect of A Better Chance, ABC, and executive director of the Educational Policy Center. His various activities have included work with the Peace Corps, and he has been a member of the faculty of the Urban Studies Department at Queens College in the City University of New York. Mr. Boyd.

BOYD: Thank you very much. Good afternoon. I want to pick up from where the last panelist left us and jump forward a little bit. I would like to have us make an assumption together for the next few minutes. Recognizing that the controversy about the validity of tests and the other terms that are in dispute is a very important one, I nevertheless would like for us to assume for a few minutes that it can be resolved. We will be assuming that a consensus can be developed about what tests can measure, about the amount of difference in test scores necessary to indicate a significant difference between individuals, and about minimum scores necessary for adequate performance in higher education. That's a tough assumption for me to make, I imagine it's a tough assumption for some of you to make, but just for the sake of pursuing a line of thought, let's make it. If we do this, if we assume that this occurs, will any problems remain in the areas on which we are trying to concentrate today? I submit that the biggest problem of all

will continue to face us. That problem involves understanding the role of test scores and other data; job criteria, grade point averages, and other things that we use in selecting students and in pursuing our employment policies in higher education.

Current mythology holds that data of the type that we are talking about, has been the basis for meritocratic systems of admissions and employment in higher education. In addition, the mythology holds that meritocracy reigned unchallenged before affirmative action. In other words, lots of people believe that the best candidate for each student or staff position always has been sought by objective and almost mechanical techniques. People who accept this meritocratic myth can, I think be divided into two groups. Most are misinformed in that they don't know, or have forgotten how non-meritocratic admissions and personnel policies always have been. The rest are bigots who know the current system with its built-in preferences. They try to perpetuate the meritocratic myth as a screen behind which to continue business as usual in order to maintain their monopolistic holds on opportunities.

What really is the situation in higher education? How are students admitted? How are staff members hired, retained and promoted? Let me propose some answers to these questions and to the bigger question about the role of tests, GPAs and other data of this type. The stated goals of selective admissions processes vary greatly but usually emphasize diversity as well as academic achievement and potential. For the moment we are assuming that academic achievement and potential can be determined, so attention can be devoted to diversity. In the admissions process consideration frequently is given to factors such as academic interest. In other words what prospective major does a student have? I think that it is very clear that if you apply to certain institutions which have three physics faculty members, all of whom have tenure, and two physics majors and you say you want to be a physics major, you are in. Extracurricular interests—if you have a marching band that doesn't have a flute player and you play the flute, you are in, unless they are ready to get rid of the marching band. Geographic origin—very obvious. It is a lot easier to get into Harvard from New Mexico than it is from Boston. Type of secondary school you attended—a great deal of attention is paid to this, trying to make sure that all of the students don't come from the same type of institution. And, certainly not last, but the last one I will mention, the existence or absence of alumni of the institution in question in the family of the candidate in question.

Now, it therefore is preposterous to say that a purely meritocratic system of admissions is about to be corrupted for the first time by expending this list of factors to include racial or ethnic background, sex or socio-economic status. The merit a student has for a particular college opening has been, and will continue to be, measured by multiple criteria no matter what happens to affirmative action.

Now, the employment process in higher education is similar to the admissions process. Committees which deal with personnel policies do not base their actions on achievement and potential as measured in various

concrete objective ways. In fact, they consider a variety of factors which occasionally, *occasionally*, include teaching ability. Another factor considered is willingness to teach a course no one else wants to cover. Having sat on the Personnel and Budget Committee in the City University, I know that one very well. That willingness to teach is not always associated with proven ability to teach. That is, somebody has got to teach that course and if you are willing to do it because you want the job, it's yours.

Another factor considered by these personnel committees is diversity, or the lack of it in the graduate institutions represented in a given department. You have a department of ten people and you have six of them who are PH.Ds from Berkeley, people start to wonder if you are doing too much of what you referred to, incestuous selection. I could go on with this list without any difficulty whatsoever. But I will stop with the list after asking you to think about one additional factor which does receive attention.

I will ask you to think about it by putting a question to you. How many times have discussions about hiring with which you are familiar, changed course after a subjective comment? Here is an example of this kind of subjective comment. "I agree with the rest of you that this candidate's resume and publications are not outstanding, but he or she made a fantastic impression on me during our interview", or "when I heard her or him speak in such and such a setting". To summarize what I am trying to say, I think all of us have to be wary of pitfalls involved with basing policies and actions on myths. If people do not learn, or remember how admissions and employment processes work, they can become outraged by developments which should be understandable and acceptable. Once one accepts current myths about admissions or employment, it is easy to accept charges that "reverse discrimination" is common-place, because Blacks, or women, or other minority groups are chosen even though they do not meet certain standards. And it's easy to do that without going back and asking yourself the antecedent question about the real applicability of those standards in conceptual terms, or the real extent to which those standards have been met by yourself or by other people who are already in, or are already on the wagon that the Blacks and women are trying to get a ride on.

I think a concluding anecdote may help make my point. It involves a fairly well-known American by the name of Henry Aaron. I am sure that many of you, certainly those from around the New York area, certainly any who are sports fan, picked up on a story which appeared in the *New York Times* Sports Section one Sunday in September,² right after the Commissioner of Baseball had said that it was about time that with all the superstars in baseball that one team came up with the guts to hire a Black manager. And the response from an also-ran white baseball player of some years, appeared full in the *Times*. I won't go through the whole response—though I recommend it to you as a very interesting bit of reading. The point is, this individual charged, and other people picked up the charge that the mere notion that Henry Aaron could be a baseball manager was indicative of the fact that, what he called "affirmative racism", had gone wild in this country, because clearly—without any doubt whatsoever—Henry Aaron was unquali-

2. N.Y. TIMES, Sept. 8, 1974, § V (Sports) at 2, col. 5.

fied by any criteria, other than that he was Black, to be a manager. It even went so far as to tell the outrageous lie that Henry Aaron had said he should be a manager just because he was Black. The charge was, and I hope you listen to this even though you may say baseball is a far cry from higher education because it doesn't sound that different to me from some of the things I hear—the charge was that Henry Aaron was “not qualified to be a baseball manager because of his deficiencies in terms of ability to read and write well”, that's a quote and if anybody wants to read it, it is here, “and to express himself lucidly”. Now of course, if that or those characteristics are essential for being a baseball manager, some very well-known people—I can think easily of Casey Stengel and Yogi Berra—obviously are woefully unqualified for their jobs. I say to you as we think about this, and laugh about it, and perhaps feel a twinge of sadness about it in 1974, that the pattern of requiring more of minority group members and women than is required of white males for the same selection process or for the same good, whether it is a classroom space or a job, that that pattern must be broken and it must be broken in higher education as well as in baseball. Thank you.

MUNTER: Well, now we have about 17 minutes for questions. May I ask that you try to address your question to a specific member of the panel and that you try to be as succinct as possible in stating your question or comment.

QUESTION: I would like to ask Mr. Enneis when the new form EEO-6 will be available and any additional information he can give us about it in a concise form.

ENNEIS: Okay. It is now scheduled to be mailed out next fall. It will not be mailed out this year.

MUNTER: Let me take the prerogative of the moderator and ask a question. This is the first time I have ever heard a discussion about any kind of testing in which the notion of privacy and confidentiality didn't arise. I wonder if any of you gentlemen on the panel want to comment on that insofar as testing procedures and even the search for ability may raise questions about personal matters. Someone mentioned that question of pervasion of the data, misuse of the data that is collected, and so on. I wonder if anyone would want to comment on either the risks that testing procedures for selection, or diagnosis for that matter, may present to affirmative action procedures or the protections that may be there. I am thinking now of course, of threatened actions that are being taken in local governments and on the federal level as to access to such records.

ENNEIS: I have some very strong private opinions about the collection of unnecessary data on individuals of a personal nature. This is not an Agency policy and that is not what the EEOC is all about. Of course data can be misused. I don't think any more data on a job applicant, or as far as that is concerned, on an applicant for admission to a school, should ever be collected than is necessary to make an intelligent decision. I think the only data that should be collected are those that have to do with job relatedness and some administrative things. For example, asking social security numbers or asking whether the individual has dependents or something of this

nature. Possibly for administrative reasons, even for tax reasons or the like, but otherwise, I just don't think it should be collected unless it is related and known to be related to job performance. I think that is it. For example, you know the polygraph is a very popular instrument in many places. Of course, there are now I guess over a dozen states that outlaw its use in any employment context. I think Massachusetts is one of those states. But you know what, I have yet to see one state, and I mean one, in the entire psychological professional literature that shows that any kind of results from polygraph, that is the so-called lie detector, have any relation to job performance in any way whatsoever and that includes honesty on the job.

LEONARD: What happens to the data, either that which is related or the extricated? What happens to it after it is collected? Who controls it? I am particularly concerned about ETS. ETS probably has the greatest collection of data next to CIA, and it is totally uncontrolled, it doesn't answer to anyone. What happens to that data? Do you sell it, do you make large studies, what happens to it?

WILLIAMS: Let me answer that question second. I will respond to this one first. Privacy is a big problem in any mass data bank and most institutions have some internal sets of rules and regulations of varying degrees of complexity as to how they deal with it. I think that the Buckley Amendment is a solution (I happen to think it is a fairly good one) where one control on misuse of those sorts of data is to insure that the individual who has provided you with information on whom you have a file, always has access to the information for correction or for request for destroying the information.

Coming back to Mr. Leonard's question, what happens to the data at ETS? I think I can respond to that, I don't know about other places. ETS, just for an historical bit, was started by essentially three organizations, the College Board, the American Council on Education and the Carnegie Corporation. Each of these organizations had a testing program. One for admission to colleges, another for admission to graduate schools and the other a teaching examination which is now referred to as the National Teachers Examination. These organizations, along with other groups whom we serve, such as the Law Schools, Graduate School Deans, Business Schools, etc., have what are called governing councils. Some of them are incorporated like the Law School Council, some are not, like the GRE and others like the College Boards, of course, are well incorporated. They have the legal last say on what happens to the data. The different programs have different rules. Some of them have a "destruct rule", after five years. Others have "no, hold it for purposes of research", of aggregate research. I am not really sure at this particular moment what the law school's rule is. In general I think there has been an interest at ETS, for a fairly simple reason, in a no longer than two-year retention rule.

The reason for that is saving data takes space and taking space means that you have an inventory problem, not only a security problem but you have the expense of just keeping it. You have to keep it in some machinery that will form the data and that has to be updated periodically just because paper cards wear and they can't be read in a machine after a while; or

magnetic tape becomes old and you have problems with that; or you might change the machinery and then you have the problem of how do you read some other media. From our vantage point, it would be desirable to destroy the data. There is an interest of course, in trying to continue validity studies. But and if—and again someone raised the question this morning—there isn't much information as to what makes a successful lawyer, what makes a successful physician, psychologist, etc., we had to know to what extent these sorts of indicators are indicative of what happens after the professional school training is over. Well, there is not much, there is not any that I know of, although there have been proposals for the same, studies that would indicate that. If there is to be an on-going requirement that tests are used in academic settings, they must be validated, and there is going to have to be some sort of retention of a variety of data. The conditions for that retention have to be scrutinized. In all probability, 2 to 3 years is about as long as they last.

BOYD: One thing that is interesting to me about the reactions concerning privacy, particularly in higher education, is that a lot of the hue and cry has been about recommendations. And I find it very interesting and disturbing that the concern is not privacy, it is whether it will botch up the process. The fact that someone knows that the person involved could read the recommendation, will make them be more careful about what they say. I think because that is cloaked under a concern for privacy but really isn't a concern for privacy, I am not terribly worried about it. I think with you, that the Buckley Amendment, strange as its source may be, is a good step and I think from experience that I have had in systems where it is made clear that the recommendation is visible, or is accessible to the person being written about, that it may make you be a little bit less colorful in your language. But if you are an honest person you can still say what you say. I think that it is one of those things where the end is predicted, I think, well before it is about to come.

WILLIAMS: There has been one implication on this—I don't know how it is going to come out as far as the court is concerned, but one fear has been that if an individual has access to all of his or her data, and a potential employer knows that, the employer might request, or order, or require, the prospective employee to present that for file. I don't see that as being implied in the Buckley Amendment, but that has been a kind of concern that some institutions have been worrying about.

COMMENT: There may be a way of getting around the problem Mr. Boyd mentioned by not writing recommendations, but by doing something different, less confidential, more superficial.

BOYD: It's like inventing a better lock. I never have been one to assume that if you invented a better lock a burgler couldn't invent a better burgler tool. That is not what I was trying to say. I was merely trying to say in those cases where the system is maintained, and I have been part of them, particularly for department recommendations on faculty, where you had a system where a person went and saw your class and wrote a recommendation, it went mysteriously into a file and then you had to wonder what happened. We had a system, I think it is all over the CUNY system in

which those are open, and it's known that they are open, and you write them and you discuss them with the colleague that you wrote them about. I think it is a step forward in terms of the *real* privacy—that is, your right to know what it is that is affecting your job opportunity and why you did or didn't get into a college and this kind of thing.

QUESTION: Mr. Williams, is it at all feasible for Educational Testing Service and other colleges, to be concerned about the fact that you have a lot of data on a lot of people, and that if you ask them they may not want you to keep it? Maybe you ought to ask them.

WILLIAMS: Is it possible that that was a concern? Yes.

QUESTION: Yes. I think ETS has always been kind of famous for its feeling of public responsibility. . .

WILLIAMS: Well, I don't know if that is what we have been famous for, but I will accept that.

QUESTION: Well, it seems to me this is something that ought to be thought about very, very carefully by the people in Princeton. You have all this data, you say you don't want it, it is taking up your space, wouldn't it be interesting to ask people if they wanted you to have it? Would there be a financial problem to do that, or is it an in-house question, or a legal question?

WILLIAMS: No, it is a jurisdictional question, but it is not a financial question.

That reminds me, I didn't really answer another part of Mr. Leonard's question. The issue as to whom does the data belong is one I suspect that really hasn't been thought about carefully. Institutions tend to claim the data as their own. But typically all the information that we have, and this is not true for all other organizations, about an individual, is provided by the individual. If it were the case that some of that information were sold and there were income generated from that, then the question, and it seems to me a legitimate question to raise, is should not the original owners of those data gain some of that income if that is what they so chose to do? The data are not sold but I do think that is still a question and I don't think the other one as to who owns the data has really been answered. So far the kind of answer that has been proffered has been that the particular governing councils own their data. We can probably defend that from the legal vantage point that we are a service agency, but the fundamental question still resides as to what do you do with this and how ought you to begin to think about it, such that both students and other clients such as institutions can begin to think about it too? Actually that is going on.

QUESTION: One thing I would want to say to you is that I don't want you to have it. Destroy it. Now, the question is do I have a right to do that? I don't mind your using it, that was my contract with you.

WILLIAMS: Well, there was a recent scientific article in the *New York Magazine* with some statements about that. There is a unanimous view yet. I happen to think that if you provide someone with a vital resource from

which they can generate some benefit then you ought to be able to share in that benefit.

COMMENT: With regard to Mr. Boyd's statement about the question of privacy and access to the records, I think there is another part to that and I think a lot of institutions are going to have to do a lot of work to get around that, or to find alternate means for the selection process for admissions. One of those is, would you compare this applicant with other students from your school who have applied and come to this school? I would say that if that kind of comparison were made, that that could be the basis for some change.

The second thing is these applications from schools which ask deans of students and other administrators if they can tell of any information which would not ordinarily come to the school's attention on the basis of the other ordinary criteria and documents being asked as part of the part of the admissions process. I do think that privacy is a concern.

BOYD: I didn't realize that statement would cause that much trouble. I don't want to have to defend it, but I think that one thing that I feel very strongly about is that it would be very good if institutions had to work damn hard to get around all these games they play now with no difficulty at all. I remember when we first got access to the records at my institution and found all of these funny circles around the ends of peoples names. Now, I think they ought to have to play that a little bit closer to the vest, if they want to do something like that they ought to have to come up with some James Bond techniques. It seems to me that it is a positive thing if you are going to say we live in an atmosphere where there is a *DeFunis* behind every stick, you know, waiting to shout, "you compared me to him and I came out better", then you have to come back to where I was in my original remarks. That is, "that's not the only way we compare you".

As long as it is comfortable to leave the myth mentioned earlier, in spite of all of the things, and I don't know if you contest it. . . . For example, that *New York Magazine* article talking about 70 points (on the math, I believe) being an inadequate difference to show that the people really were not basically comparable. Now you are going to come and sue me for not letting you in because you have 10 points more than somebody else, and 70 points is insignificant. I think you wouldn't have that kind of law suit, you know, an expensive, difficult process, you wouldn't be so anxious to run out and do that if the institution came clean up front and said that is not the only thing we compared, number one. And number two, when we compare it we have to compare it within these parameters because 10 points means nothing, absolutely nothing. And I think that is the kind of measure of difference we are talking about in the *DeFunis* case.

WILLIAMS: Could I just make one comment—from my Peace Corps experience also? One of the things that we certainly did in the assessment function was to share the entire folder that had been accumulated on each trainee. That happened to have very beneficial effects. It meant that the psychologists involved had to justify and to communicate why he or she was making the decision, whatever that one happened to be. I always worried about that. There were field assessment officers and then there were field

selection officers and when in doubt the field selection officer always had the final word, but he also had civil service background check information. If you ever became a field selection officer, then you knew that in that particular folder you had FBI contacts, etc. One of the procedures was to go to your initial set of references and ask that set of references to refer other people who knew you and on down the line. Then things went into your folder. Some of which had no meaning in terms of whether this person said he broke the window one day or whatever the case might be, and that's an indication of impulse control. I happen to think that all those data should be shared as a significant control.

MUNTER: I would like to thank our panel members for sharing their experience and wisdom with us and especially their candor. Thank all of you. Good afternoon.