THE SHUFORD-MASSENGILL CORPORATION

INDIVIDUAL AND SOCIAL JUSTICE IN OBJECTIVE TESTING

Emir H. Shuford, Jr. and H. Edward Massengill, Jr.

BLANK PAGE

INDIVIDUAL AND SOCIAL JUSTICE IN OBJECTIVE TESTING*

Emir H. Shuford, Jr. and H. Edward Massengill, Jr.

Throughout the history of objective testing concern has been expressed about the effect of guessing on test scores. This concern has been intermittant with some investigators finding that guessing is a terrible problem, while others have attempted to show that guessing is not so bad after all. Textbooks in the area of psychometrics and educational measurement tend to deal quite briefly with the effect of guessing and point to the conclusion that guessing poses no real difficulty for applications of objective testing.

This relative lack of interest in the problem of guessing is really not a surprising phenomena when one remembers that there has been no satisfactory alternative to objective testing. Now, this has all been changed. The recent application of the logic and techniques of decision theory to objective testing has resulted not only in an improved understanding of the nature of guessing (Massengill & Shuford, 1967), but more importantly, in the development of new methods of objective testing which have all of the significant advantages and none of the disadvantages of the older forms of objective testing. In particular, the new method of Valid Confidence Testing, based on an admissible scoring system (Shuford, Albert & Massengill, 1966), completely eliminates the problem of guessing.

Thus, it is no longer an academic exercise to investigate what effect guessing may have upon the data from objective testing and the decisions based upon these data. In fact, such studies can provide useful guidance in the decision to change over to the new and improved form of objective testing. While these theoretical studies cannot estimate fully all the benefits that might flow from Valid Confidence Testing, they can give some insight into the price that is being paid by continuing to use the older forms of objective testing.

Logical and mathematical analyses lead one to conclude that this price is not small. For example, both the reliability and validity of choice testing is severely degraded by the existence of guessing (Shuford & Massengill, 1966a; Shuford, 1967). Guessing can cause severe losses in the performance of selection, classification and placement programs (Shuford & Massengill, 1966a). Guessing can so distort test results that in some instances it is best not to use ability test scores for counseling purposes (Shuford & Massengill, 1966a). The existence of guessing on classroom tests places an unacceptable limit on the effectiveness of instruction (Shuford & Massengill, 1966b; 1967). Finally, testwiseness can be the dominant factor in determining who passes and who fails a test (Shuford & Massengill, 1966a).

Now, let's go into this last result more deeply. Figures 1-4 have been adapted from Figure 9 in the section on testwiseness in *The effect of guessing on the quality of personnel and counseling decisions* (Shuford & Massengill, 1966a). These results apply to a ten-item true-false, multiple choice, or fill-in-the-blank test. In other words, they apply to any ten-item objective test.

The mathematical analysis is based on the assumption that each individual has an ability or achievement level which can be characterized as the proportion of test items to which the individual knows the answer. An individual knows the answer to a test item if he is able to express a consistent and stable preference for the correct answer to that item. Ability level in terms of the proportion of items known can range from a value of zero up to a maximum value

*The third semiannual technical report (which covers the period May 1967 through October 1967) of work performed under contract number AF 49(638) - 1744, ARPA order number 833, by The Shuford-Massengill Corporation, P. O. Box 26, Lexington, Massachusetts, 02173.

of one as shown along the abcissa of Figures 1-4.

A particular test is given as a random sample of items from a pool of test items covering the relevant body of knowledge or ability area. Thus, if one knows the ability level of a person. it is possible to compute (by using the binomial distribution) the probability of the person making any particular test score. These probabilities may be cumulated to find the probability of the person making a score high enough to equal or exceed the cut-off score and pass the test. These probabilities are shown along the ordinate in Figures 1-4.

The dotted line shows the probability of a person of any given ability level passing a ten-item test if conditions are such that no guessing occurs on the test. If a person adopted the test-taking strategy of skipping those items on which he would have to guess, then guessing would be eliminated as a factor in his test performance and his chances of passing the test and being selected into some program would be as shown in Figures 1-4 for cutting score values of three, five, seven and nine.

Now consider what happens if a person chooses not to skip those items for which he would have to guess at the answer. By going ahead to attempt these items, the person would get a certain number of them correct by chance and the highest possible level of chance success is, of course, one-half. A probability of chance success of one-half must characterize the situation for a true-false test. This should be self-evident. It may not be so clear, however, that the same maximum level of chance success may obtain for four-and five-alternative multiple-choice tests and for fill-in-the-blank type tests. This can happen when the person has enough information at hand to exclude all but two of the possible answers. In this case, the person goes ahead and guesses at which of the two answers are correct with the result that the probability of chance success is at the maximum level of one-half. The dashed line in Figures 1-4 represents the probability of passing the ten-item test for a person of any given ability level who is guessing with probability of chance success one-half.

It should be clear from an examination of these figures that whether or not a person chooses to guess in taking a test makes a great deal of difference to his chances of passing the test. In many instances, if the person refuses to guess, he stands almost no chance of passing the test. while if he goes ahead to adopt the guessing strategy, he is almost certain to pass the test.

Now, some might think that these results are not fair to objective testing. Any responsible producer of tests is aware of the effect of guessing and the serious consequences of the decisions taken on the bases of the results of his tests. Certainly, no responsible test producer would advocate the use of a ten-item test for selection, classification, placement or counseling purposes. Most tests are much longer than this.

So let's see what happens for a 100-item test. The relevant computations have been performed and are plotted in Figures 5-8. Examination of these figures indicates the decision to guess or not to guess has even greater consequences than before though over a narrower range of ability levels. Is this an improvement? It may be from the point of view of the user, but it may not be from the point of view of the person taking the test.

But maybe this is just another academic exercise since all responsible producers of tests have adopted the policy of instructing everyone either to guess or not to guess. So, if everyone follows these test instructions, which are the same to all people taking the test, then the test is fair in its treatment of each individual.

Does everyone follow the test instructions? When the instructions say to guess, does everyone taking the test immediately adopt a guessing strategy? Guessing seems to run against the grain of some people. They don't like to guess; they have trouble guessing. In a guessing situation, the person has no information available on which to base a preference for one of the answers. He struggles trying to recall information to decide the issue. Some people hate to resolve it by a flip of a mental coin. Maybe they don't like to appear the fool when they guess the wrong answer.

Some people tend to be this way. There is not any doubt that the roughness of the grain is sanded down by many years of taking objective tests in the schools. And more particularly, in the better schools at the higher levels of education, the finish gets pretty smooth and one has no difficulty at all guessing when taking a test.

Those instances in which the test instructions imply that guessing hurts one's score and that the person should not guess are especially noteworthy. The interest resides not so much in the fact that the instructions are based on a lie, but upon the possibility that some people have been afforded the opportunity to learn that it is a lie, while others have not.

We have examined many scoring systems that have been used over the years, and we have yet to find one that penalizes guessing in such a way as to eliminate the effects of guessing on test performance (Massengill & Shuford, 1966; 1967). In particular, take the "correction for guessing" formula used for many multiple choice tests. It in no sense penalizes guessing. At most, it eliminates the advantage from guessing when the person is maximumly uncertain among all the possible alternatives. If the person has enough information at hand which disinclines him toward one or more of the alternatives, then it is definitely to his advantage to guess. It is to his advantage in the sense that his expected test score would be increased by so doing. Incidentally, the "correction for guessing" scoring formula is not even a correction formula for guessing in any real sense of the term (Massengill & Shuford, 1966).

As far as we can tell, the responsible producers of tests who instruct persons taking their tests not to guess are not aware that they are deceiving these people. Their psychometricians apparently applied some kind of intuition to analyze the situation and concluded that their scoring systems penalize guessing. In fairness, it must be said that most of these conclusions were reached many years ago before the logic of decision making had become very widely known. Now, however, it is a simple matter to analyze any proposed scoring system and to advise a person taking a test as to what test-taking strategy to use in order to maximize his expected test score (Massengill & Shuford, 1966; 1967; Shuford, 1965).

Now consider the situation of a person who has just been instructed not to guess on the test that he is about to take. This person might yield to the voice of authority and skip those items to which he does not know the answer. If, on the other hand, this person has been to the right schools and as a consequence has taken many objective examinations, he has had the opportunity to try out different test-taking strategies and to discuss it with his peers. In this manner he can learn that it does, in fact, pay to guess on all objective examinations. He can know through experience that it pays to guess no matter what the test instructions say about it.

So, we have some persons believing the instructions and skipping those questions to which they do not know the answer, while others know better than to believe the instructions and go ahead and guess at all the test items. As we have seen, Figures 1-8 give some indication of how these persons' chances of passing the test will be affected by their decision to guess or not to guess.

A particular person's chances of passing the test is a matter of fairness to the individual. When we consider, however, how groups of people make out on the test, it can be a matter of social justice and one with some practical consequences.

We have suggested that with the right educational or cultural opportunities a person can learn that it is in his best interest to always guess in objective examinations. If, on the other hand, a person has never had these experiences then he may be reluctant to guess or he may be inclined to follow the invalid test instructions not to guess. When this happens, it can be a source of cultural bias that is at least as important as the more well-known kind.

Much of the discussion of cultural bias in testing has focused upon the content of the test, that is, what questions are asked and the language used in asking the questions. For example, a test item may inquire about an object or activity which is very familiar to one class of people but which is outside the common experience of another class or the language or terminology used in the test may be natural to one class but unfamiliar to another. In both instances, interpretation of test results becomes more uncertain because we can never be certain why a person failed an item. A person failing a test item in spite of having many opportunities to become familiar with the knowledge tapped by the item is quite different from the person who failed the test item because he had no exposure to the information. In a few instances, this difference is not important. For example, it may be that a person must have the knowledge tapped by the test in order to perform a job. More frequently, however, the difference is important. Conditions may be such that a person may be allowed to acquire the necessary information before performing the job. Or, more generally, in any case in which the information tapped by the test is not centrally related to the behavior to be predicted it can make a great deal of difference whether or not a person has had the opportunity to learn this information. The most striking example of this is probably in the area of ability testing where validity rests on the assumption that all people have an equal opportunity to learn the material covered by the test. It is just this assumption that makes an achievement test into an ability test.

Now this type of cultural bias is widely known and commonly understood. The new kind of cultural bias which we suggest may exist does not rest upon the content or language of the test items but instead is an outcome of the very process of testing. No amount of content revision and/or item rewriting can eliminate it. This new type of cultural bias resides in the method of test administration and can only be eliminated by changing these methods.

If cultural bias based on differences in test-taking strategy does exist and even if it exists as a very strong factor, it would be very hard to find it in the data of objective testing since it would occur whenever an objective test were used. This means that it would be independent of the content of the test and it would appear, for example, as a common factor in all factor-analytic studies using only objective tests. This general factor, with loadings on all tests, would appear to be the general ability factor and might be abbreviated as the G-factor. The G would be correct if one interprets it to be guessing and not general ability.

Now, how important is this source of cultural bias? We have already seen through the use of logic and mathematics that guessing or not guessing can make a great difference in a person's chances of passing a test. This finding suggests that differences may also arise when we consider groups of people who guess vs. groups of people who do not guess on an objective examination. If these differences appear to be important, then we may conclude that guessing on objective examinations may be a significant source of cultural bias. It would be an important source of bias to the extent that it is associated with certain groups of the population who have has different sets of experiences and opportunities to learn effective test-taking strategies. To the extent that the analyses indicate that differences in the test-taking strategies of different groups yield only minor differences in test performance we can be less concerned about guessing as a source of cultural bias.

Suppose that the population of people taking a ten-item objective examination is made up half of people who are guessing with probability of chance success of 1/2 (call these people Type A) and half of people who do not guess on the examination (call these people Type B). Suppose further that both groups have exactly the same distribution of ability levels, say, the distribution shown as Figure 1 in Shuford & Massengill (1966a). This distribution is bell-shaped and is symmetric around an ability level of .5. Thus, half of each group has above average ability level and half have below average ability level. But remember that the distributions are the same for both Type A and Type B individuals. The only difference between these groups lies in the fact that Type A people guess while Type B people don't.

Now suppose that the people in our population take the ten-item objective examination. What would be the results? One way that we can look at the situation is to find for each particular test score what percentage of the people making that score are Type A and what percentage are Type B. If the test results were just reflecting the persons' ability levels, then for any test score half of the people should be Type A and half should be Type B. The test is biased to the extent that the proportions deviate from this fifty-fifty split.

Figure 9 shows the proportion of people making each test score who are of Type B. Notice that these people dominate the groups for the lower test scores, while they are under-represented for the higher test scores. Though Type A and Type B people have exactly the same distribution of ability levels, the test results make it appear that the Type B people are lower in ability than the Type A people.

Figure 10 shows similar results for the case in which the Type A people are guessing but with a probability of chance success of only 1/5, the practical minimum level of chance success. Here the bias is much less but it still exists in a not insignificant amount.

What happens if only, say, 10 percent of the population are culturally deprived in the sense that they have not had the opportunity to learn to always guess on an objective examination. Here, the Type B people should make up 10 percent of those persons making any particular test score while the Type A people will, of course, make up 90 percent. To the extent that the test is biased, the proportions will deviate from these fair values. Figure 11 shows the proportion of people making each test score who are of Type B for the case in which the guessing level is maximum at a probability of chance success of 1/2. Figure 12 shows similar results when the probability of chance success is minimum at 1/5. Obviously the bias is still there in the test with Type B people seldom being found among the high scorers on such tests.

The results shown in Figures 9-12 indicate that this one source of cultural bias alone is sufficient to make it appear that Type B people as a group tend to have lesser ability than do Type A people. Furthermore, the assumptions of the derivation are such that it is clear that this phenomena is an artifact of the testing method. It has no relevance to the real ability levels of the peoples involved.

There is another interesting way of looking at the data: a way that has some implications for the selective employment of peoples of Type A and of Type B. Of those people who make a particular test score, some will be of Type A and some will be of Type B. Now, what is the average ability level of those people of Type A and what is the average ability level of those people of Type B? If there were no bias in the testing method, they would, of course, have the same ability levels and the ability levels would be higher for the higher test scores. Figure 13 shows the average ability level for each of the two groups of people when Type A people are guessing with probability of chance success of 1/2 and Type B people are not guessing. For all test scores except zero, the Type B people have a higher ability level than do the Type A. Figure 14 shows similar results for the case in which the Type A people are guessing with a probability of chance success of 1/5.

As before, for all test scores except zero, the Type B people have higher ability levels than do the Type A people. Thus, if people were classified into groups on the basis of their test scores, the Type B people would in general, find themselves mixed in with Type A people of lesser ability. Likewise, if an employer were offered a choice between hiring a Type A person and a Type B person both of whom have made the same test score and this employer had some means of determining whether a person were a Type A person or a Type B, then he would be well advised to select the Type B person for the job since the Type B person would be more able and thus more likely to perform better on the job.

And finally there is one more interesting way of looking at the data. It is the familiar way of looking at the performance of a test according to how well it is able to discriminate between two groups of people.

First, suppose that no one is guessing in taking a ten-item objective examination and that the people taking the test have the same distribution of ability levels that we used before in the computations. Suppose further, that we wished to use a test to separate the group taking the test into two sub-groups, those who have above average ability and those who have below average ability. Remember that fifty percent have above average ability and fifty percent have below average ability.

The results are shown in Figure 15 for each of the 11 possible test scores that might be used as passing scores. For example, a passing score of 5 means that all those people scoring 5 or more will be classified as above average in ability, while those scoring 4 or less will be classified as below average in ability. The discrimination is, of course, better for some passing scores than for others and, in fact, there are two passing scores, 5 and 6, which yield optimal discrimination with a probability of error of .24. Twenty-four percent error corresponds to 76 percent correct classifications yielded by the test when using the optimal cutting score. This performance is of some value, but remember that even without testing we could have correctly classified 50 percent of the people by just saying that all were above average or all were below average.

Now look what happens when we have a group of people all of whom are guessing with a probability of chance success of 1/2. Figure 16 shows these results. Here we see that the test suffers a loss in discriminating power with the best error rate being slightly better than 32 percent. That is, only about 68 percent of the people can be correctly classified through using the test, a gain of only 18 percent over what could be done without testing.

Now, what happens when the group being tested is made up of the two types of people, half who were Type A and guessing and half who were Type B and not guessing? How well can the test discriminate people according to their ability levels? The results are shown in Figure 17 where we find that the best error rate is about 32-1/2 percent or conversely about 67-1/2 percent of the people are being correctly assigned. This is not very good performance but it might be of value in some applications.

Now there is a different way of looking at the results for this case where half of the people guess and half don't. Instead of considering the test as a test of ability and evaluating its ability to discriminate according to ability level let's consider the test as a test of cultural background and consider its ability to discriminate people according to whether they are Type A persons and guessing or Type B and not guessing. Again, half of the people are Type A and half of them are Type B. If a person makes a passing score or better we say he is guessing and is a Type A. If a person doesn't make a passing score, we say he is not guessing and is a Type B. These results are shown in Figure 18 in a manner exactly like those evaluations of test performance when the test is considered as a test of ability. Here, we run into the rather shocking result that the test is a better test of cultural background than it is of ability level. To see this, look at the percent error column where the best discrimination is obtained from a passing score of 7. The percent error is about 26 and the percent of correct classifications is about 74. This is considerably better performance than the results obtained from the exactly analogous case shown in Figure 17.

Although psychometricians have typically evaluated the performance of tests according to the ability of a test to discriminate between people on the basis of ability and achievement levels it is just as feasible to evaluate the performance of a test in terms of its ability to discriminate those who guess from those who don't. The psychometric techniques are analogous and rest on the same foundations. We have done this for ten-item objective examinations and find that the test is a better test of test-taking strategy than it is of ability level. This result is shown in a different way in Figure 19. The test makes more correct classifications according to test-taking strategy than it does according to above average or below average ability levels.

We have performed similar computations for the case where only 10 percent of the tested population are of Type B and do not guess. The remaining 90 percent are of Type A and guess with the maximum probability of chance success of 1/2. In order to keep the base rates equivalent, the ability test was used to select out the top 10 percent of ability levels. Notice that with these base rates of ten and ninety we can assign 90 percent of the people correctly without testing by saying in the case of ability testing that no one's ability is in the top 10 percent or in the case of testing for test-taking strategy that everyone is guessing. With these base rates, there is no passing score that allows the ability test to perform better than could be done without testing at all. A ten-item objective examination used with this distribution of ability levels and with these base rates is absolutely worthless as an ability test. Notice, however, that the same test does have some marginal value as a test of cultural background or test-taking strategy since there are some passing scores which allow one to classify correctly slightly more than 90 percent of the population.

Now what do these results mean? Do they have any value other than as a shock treatment applied to responsible producers of tests? Well, first, the results are limited in generality. The numbers do not apply to all testing situations. For example, if the probability of chance success were less than the maximum value of 1/2 then the test would not be able to discriminate people according to test-taking strategy so well and, in fact, might do a slightly better job discriminating on the basis of ability level. Although this reduces the shock value somewhat, it is still true that any objective examination is discriminating against those people who for one reason or another choose not to guess when taking the test. This discrimination is unfair in the sense that it has nothing to do with the ability or achievement level per se of the people being discriminated against. It has nothing to do with their capacity for learning to perform a job or for continuing their education. At this level it remains a matter of individual fairness and of the efficiency and effectiveness of testing.

When, on the other hand, there exist groups of people to whom society has denied the opportunities to learn effective test-taking strategies, the problem of guessing on objective examinations becomes a matter of social justice. It becomes so because these groups are discriminated against not on the basis of ability or achievement but rather on their lack of opportunity, i.e., the failure to have sufficient exposure to the practices of objective testing to have learned that they should always guess when taking an objective examination.

Are there groups like this? Is this really a cultural difference? Do the educationally disadvantaged tend not to guess when taking objective tests? We don't really know; maybe it is not a factor. We assume there are very few skipped items on the College Boards indicating that almost everybody guessed at all the items. On the other hand maybe this doesn't mean too much because we doubt that many of the educationally disadvantaged take the College Boards. They have already been selected out of the system by objective testing. Objective tests are used so extensively and in such a way that if there were any such cultural bias working against people who have not learned to guess, we would have another instance of the self-fulfilling prophecy because test performance denies further educational opportunity to these people. It is then just a matter of time until their level of educational achievement falls behind those of other groups.

There is some reason to believe that guessing may be a cultural factor. Knowing to always guess on an objective examination even when you are instructed not to do so seems to require either a great deal of experience or a high level of sophistication. As mentioned before, the intuitions of most testing specialists have been insufficient to allow them to gain insight into the true nature of the scoring systems that they devise and put into practice. Many of them apparently are convinced that it is best not to guess when certain of their scoring systems are used. It takes an explicit application of logic to show that in fact this is not the case. Thus, it is doubtful that many of the people taking objective examinations have done the necessary logical analysis of the situation. Experience on the other hand is a great teacher and people who have taken hundreds of objective examinations have had ample opportunity to discover that their test scores tend to be better when they guess. So the issue can become one of experience and opportunity to gain this experience. The graduate student has most certainly had more experience of this sort than the first grader. The person who is just graduating from the pre-college program of an outstanding suburban high school has almost certainly had more experience taking objective examinations of all sorts than has the person of the same age, and possibly of the same ability, but who dropped out of a ghetto school during the seventh grade. Thus, it should not be surprising that the high school graduate is much better able and prepared to play the test-taking game than is the ghetto school dropout.

So what can be done to remove all possibility of such a cultural bias operating in the taking of objective tests? Well, instructing people to always guess on objective examinations might help, but the increased amount of guessing that would result also means that there is a lot more random error in the test scores. Tnus, the tests are even less reliable and valid than before.

Another alternative and one that is superior in every respect is to change over to Valid Confidence Testing. Since no choices are required in responding to a Valid Confidence Test, the problem of guessing per se disappears. The person taking a Valid Confidence Test can reflect accurately his state of knowledge, whatever it may be. If a person is undecided between two or more alternatives, he indicates this. He is not compelled either to choose an answer or to skip the item.

A guessing-like phenomena can, however, creep back into Valid Confidence Testing. A person can pretend knowledge by placing all of his confidence on one of the possible answers when he is, in fact, uncertain between two or more of the possible answers. In the event that the person does this and happens to place all of his confidence on the wrong answers, he makes the worst possible score for that item. If he happens to place all of his confidence on the right answer, he makes a very good score but the scoring system is such that if the person repeatedly pretends knowledge, he will almost certainly wind up with an exceedingly poor test score. In this instance, a guessing-like strategy is heavily penalized by Valid Confidence Testing. The scoring system used in Valid Confidence Testing is such that it is in the best interest of the person taking the test to honestly express his degrees of confidence, whatever they may be.

Materials currently available for the administration of Valid Confidence Tests have proved highly effective in introducing the concepts, techniques, and implications of this new type of testing to young children. Presumably there would be no great difficulties in using the method of Valid Confidence Testing with older students, military personnel, job applicants, and any others subject to objective testing. Since these materials include physical representations of degree of confidence, item score, the relations between them and consequences of various test-taking strategies and do not depend upon abstract reasoning or previous test-taking experience, everyone taking a Valid Confidence Test for the first time is on more nearly equal footing. If, however, the introduction to Valid Confidence Testing is too brief and superficial, some persons would not grasp the techniques as well as others and will not be able to respond to the test. This, of course, can be quickly discovered by monitoring the group taking the test.

It may be found through the use of Valid Confidence Testing that some people are able to evaluate information better than others. For example, it may be found that some people cannot discriminate between those situations where the information they have at hand at the moment of taking the test is sufficient to justify a high degree of confidence from those situations where there is very little and poor information available and is of such quality to justify only a much smaller degree of confidence in an answer. Furthermore, it may be found that this ability to evaluate information is related to cultural background. In this event, one might say that we have a methodological bias in Valid Confidence Testing.

We would argue that the answer is no. If a person does not evaluate information effectively, he should lack the ability to do so whether he is in a testing situation or any other situation, in school, on the job, or anywhere else. In this sense, his poor test performance would be doing nothing more than reflecting poor performance in many situations. So there is nothing about the testing situation which is unfair.

This is not to say, however, that the whole situation is completely fair. Certainly, if society denies certain groups the opportunity and experiences necessary to learn to evaluate information correctly, and, thus, to perform well at whatever they may wish to do, then we have a case of social injustice. The results of Valid Confidence Testing would, of course, accurately reflect such a situation.

REFERENCES

Massengill, H. E. & Shuford, E. H. (1966) *Decision-theoretic psychometrics: a logical analysis of guessing.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

Massengill, H. E. & Shuford, E. H. (1967) *What pupils and teachers should know about guessing.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

Shuford, E. H. (1965) *Cybernetic testing.* ESD-TR-65-467, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Shuford, E. H. (1967) *How to shorten a test and increase its reliability and validity.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

Shuford, E. H., Albert, A. & Massengill, H. E. (1966) Admissible probability measurement procedures. *Psychometrika*, 31.

Shuford, E. H. & Massengill, H. E. (1966a) *Decision-theoretic psychometrics: the effect of guessing on the quality of personnel and counseling decisions.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

Shuford, E. H. & Massengill, H. E. (1966b) *Decision-theoretic psychometrics: the worth of individualizing instruction.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

Shuford, E. H. & Massengill, H. E. (1967) *The relative effectiveness of five instructional strategies.* Lexington, Massachusetts: The Shuford-Massengill Corporation.

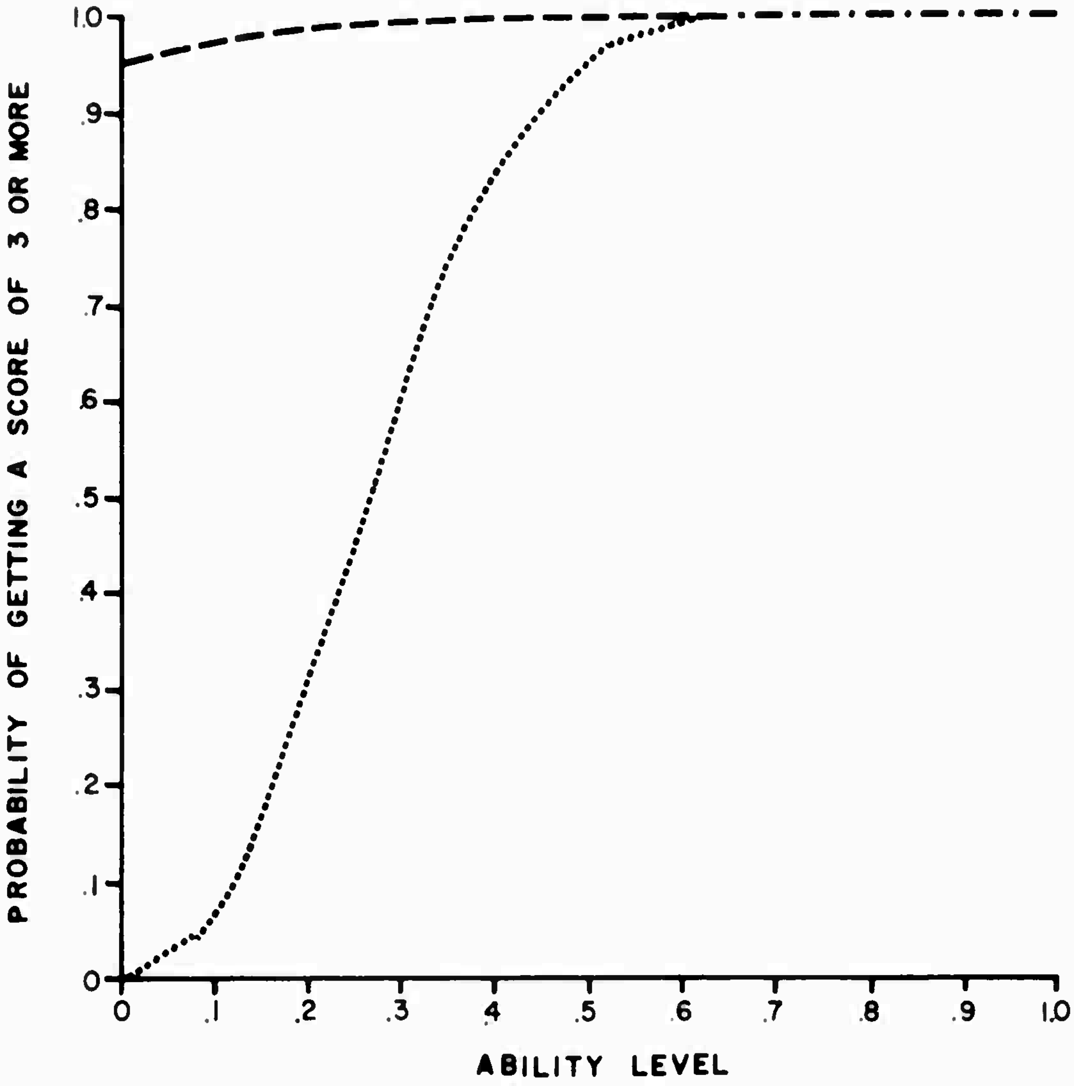

Figure 1. The effect of guessing strategy on a person's chances of passing a ten-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

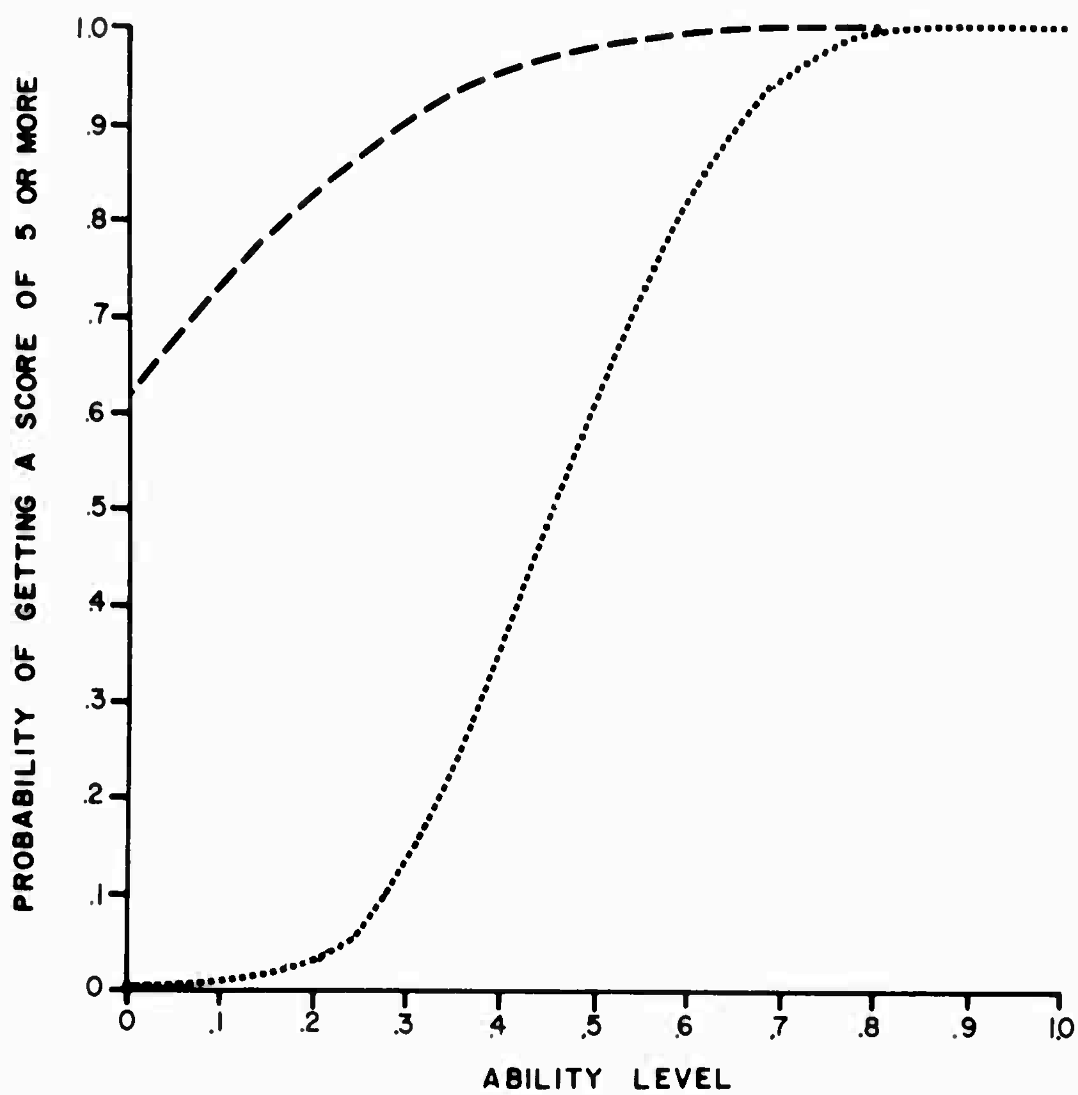

Figure 2. The effect of guessing strategy on a person's chances of passing a ten-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

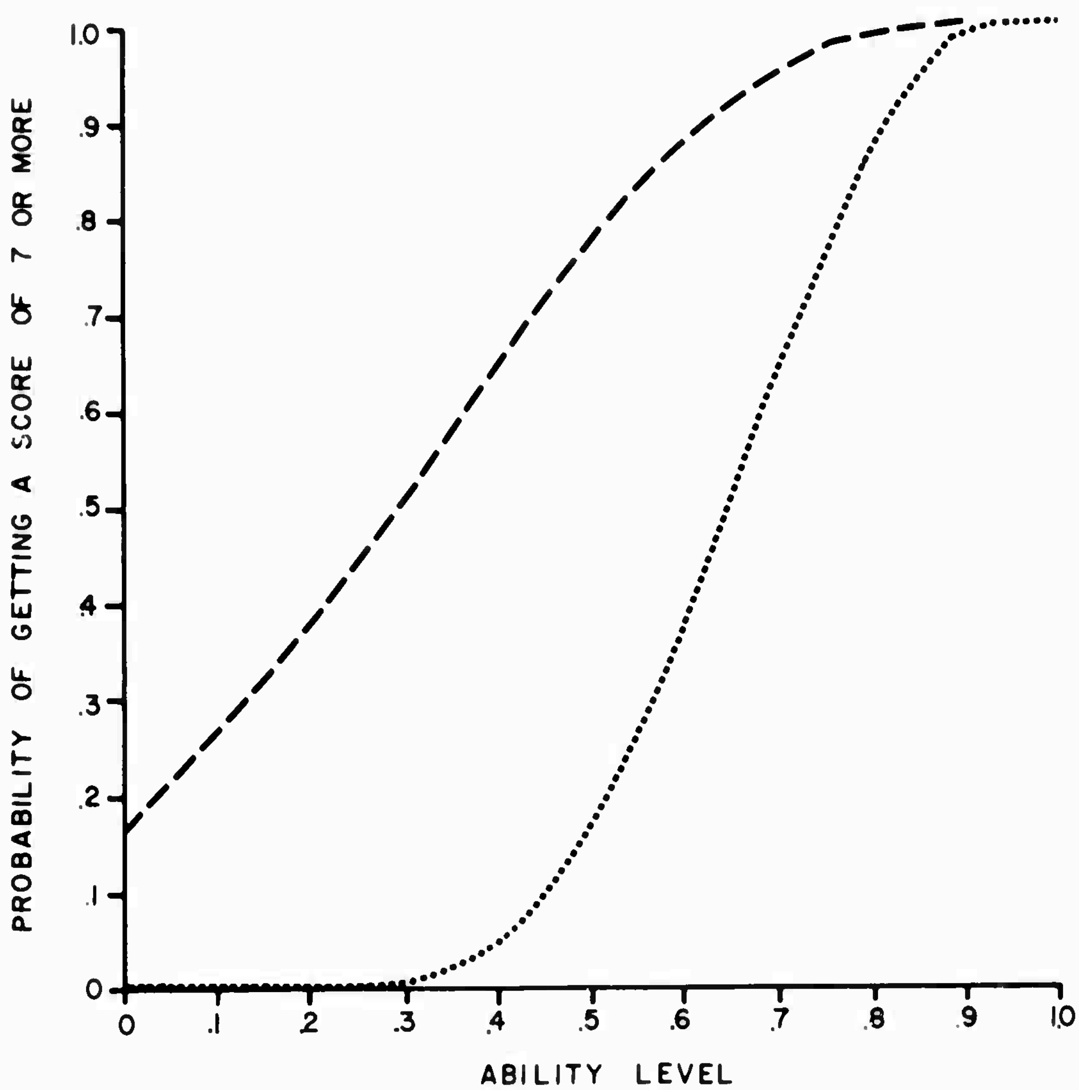

Figure 3. The effect of guessing strategy on a person's chances of passing a ten-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

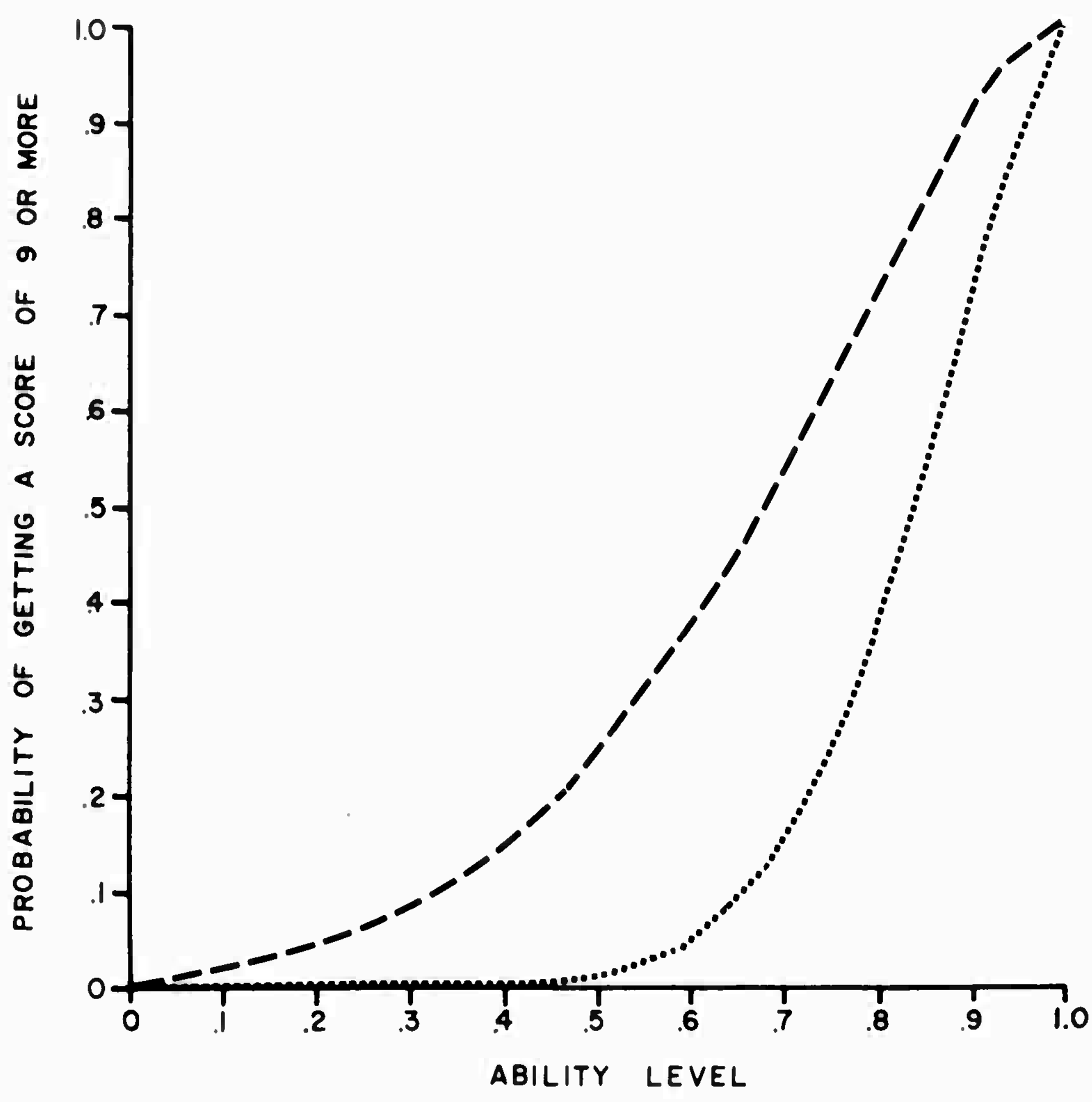

Figure 4. The effect of guessing strategy on a person's chances of passing a ten-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

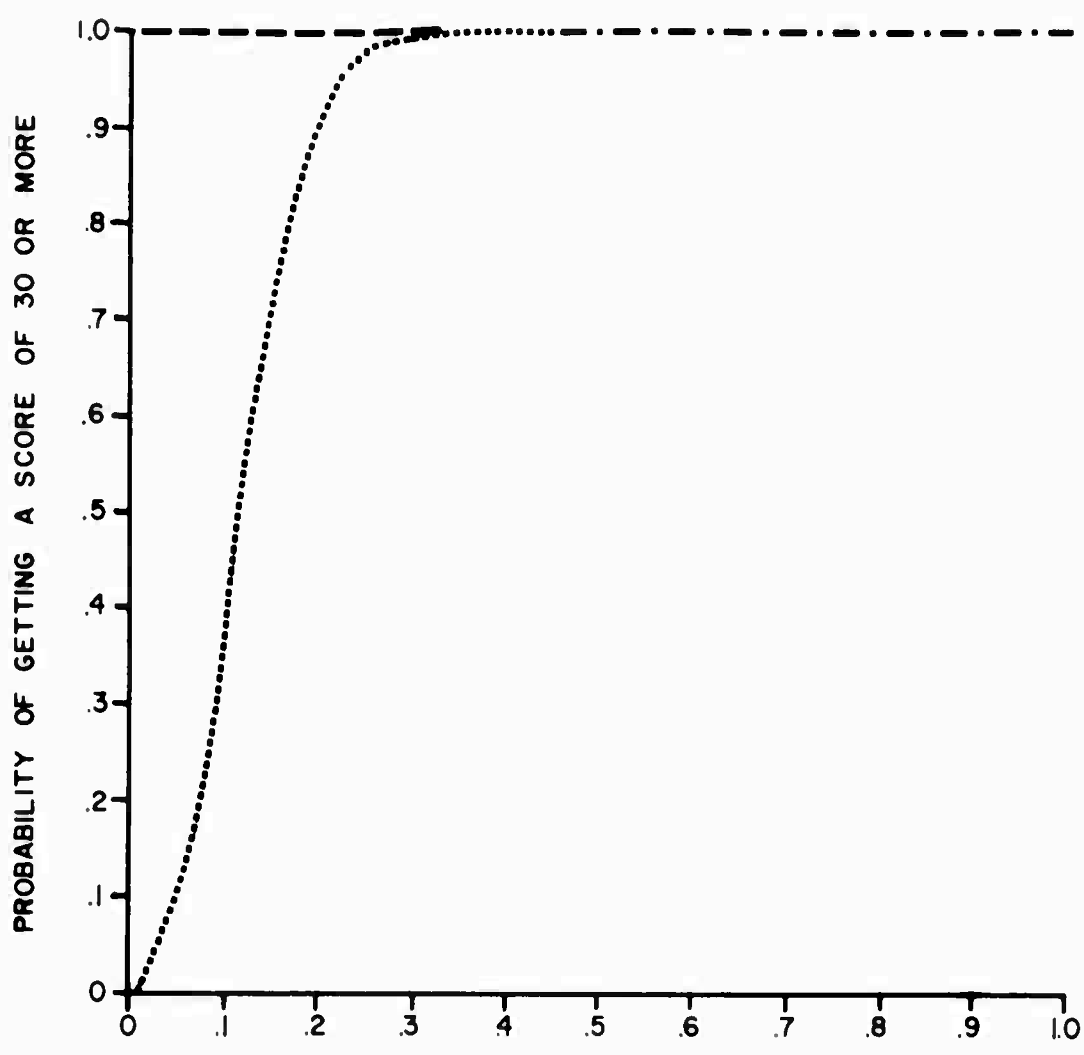

Figure 5. The effect of guessing strategy on a person's chances of passing a 100-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

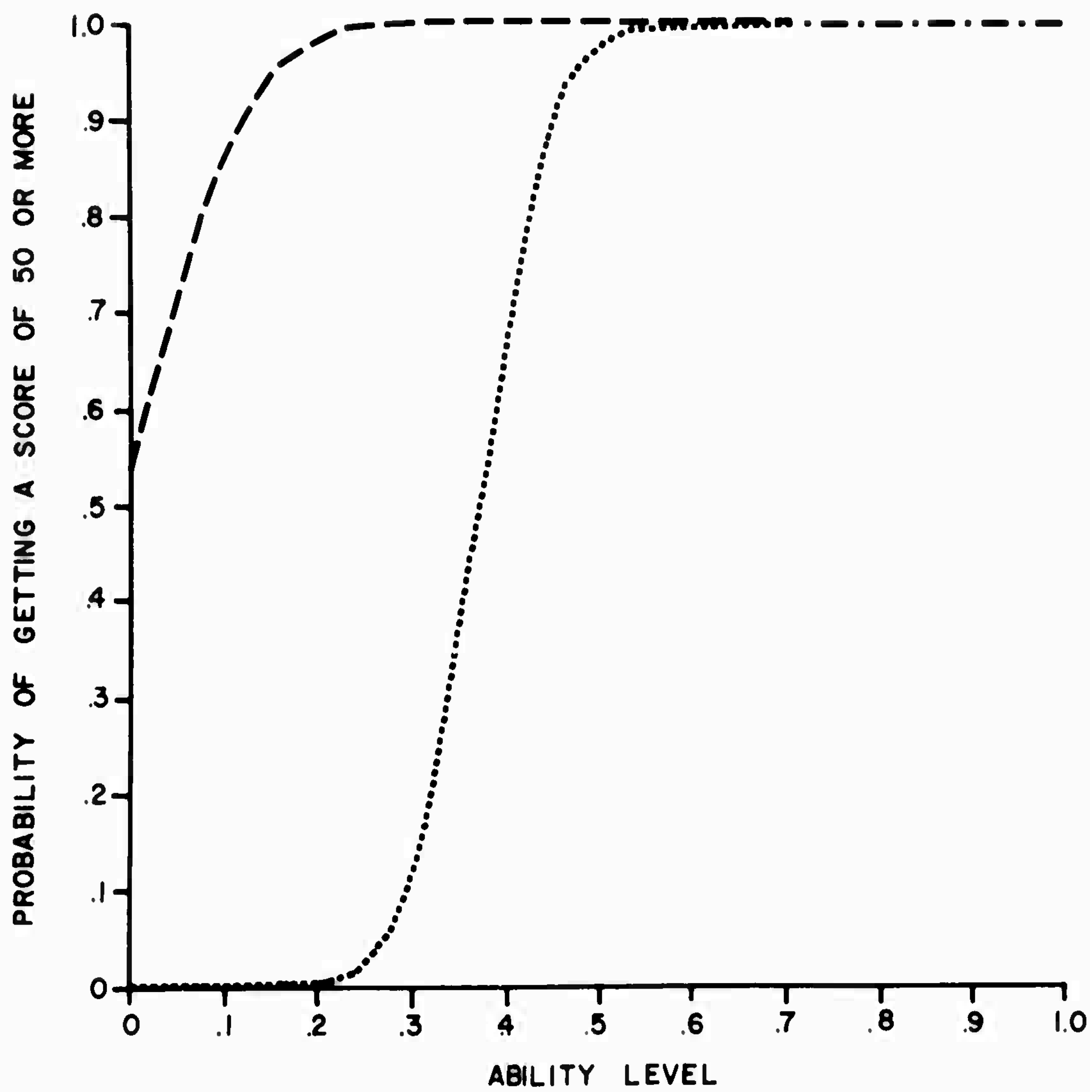

Figure 6. The effect of guessing strategy on a person's chances of passing a 100-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

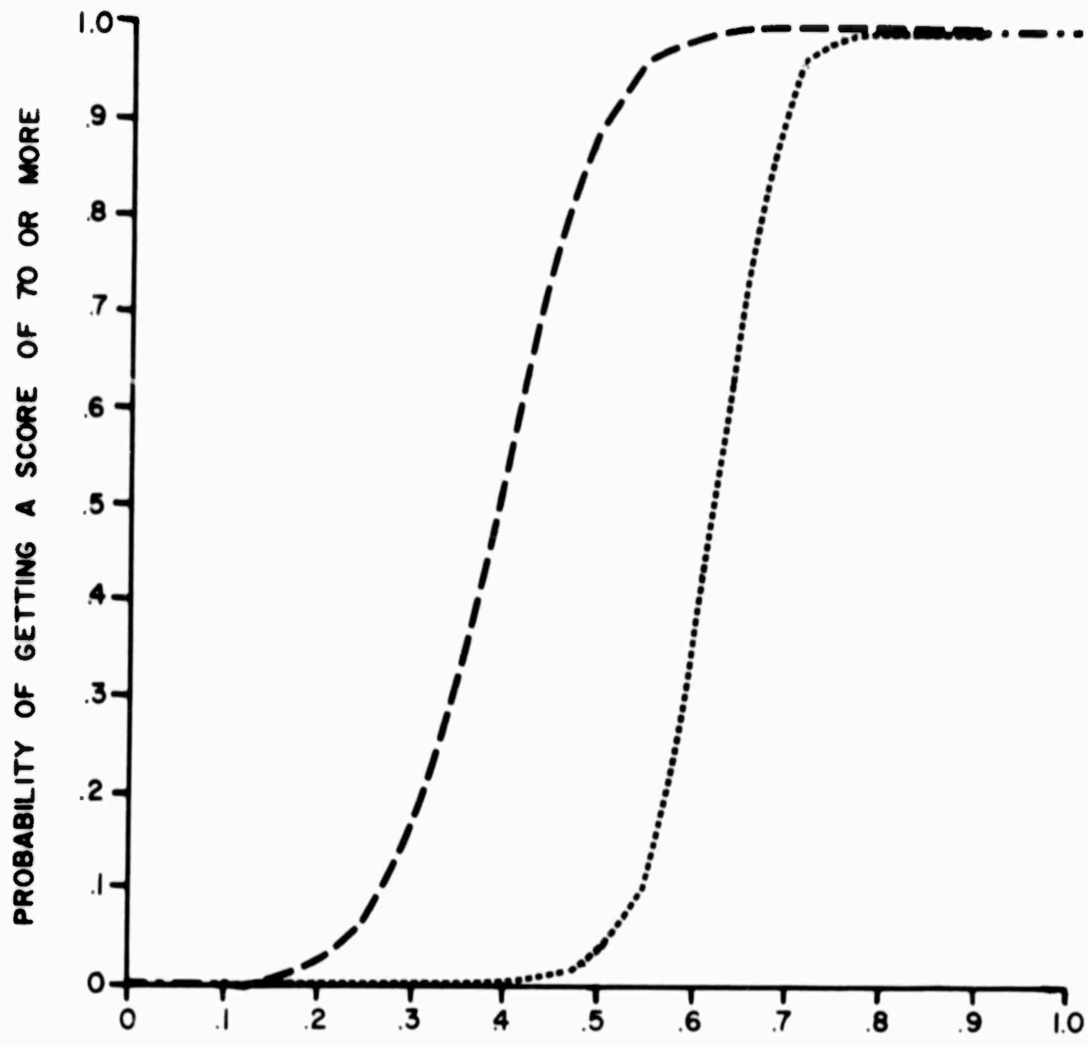

Figure 7. The effect of guessing strategy on a person's chances of passing a 100-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

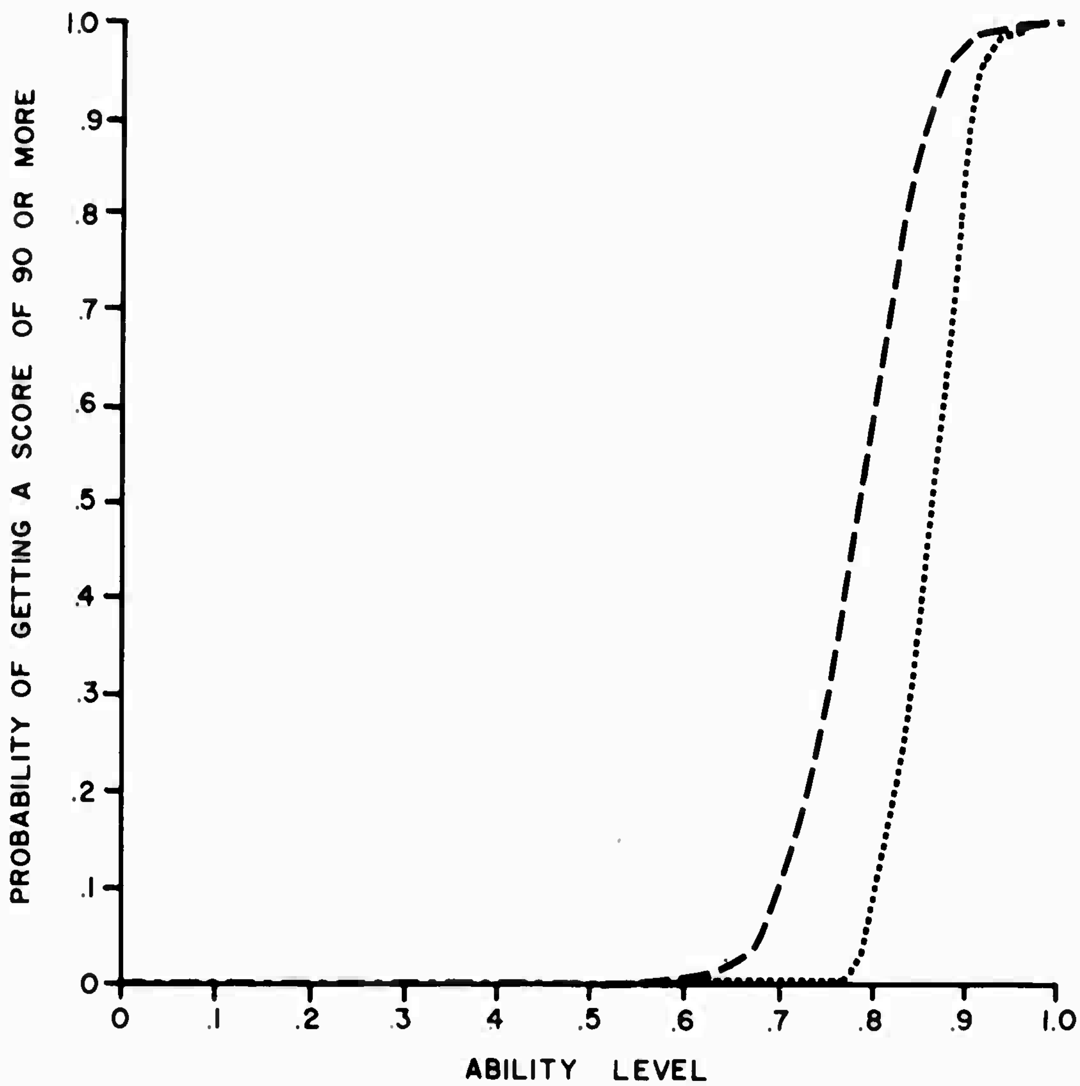

Figure 8. The effect of guessing strategy on a person's chances of passing a 100-item objective test. Dashed line: Person guesses with probability of chance success of ½ when he doesn't know the answer to an item. Dotted line: Person doesn't guess.

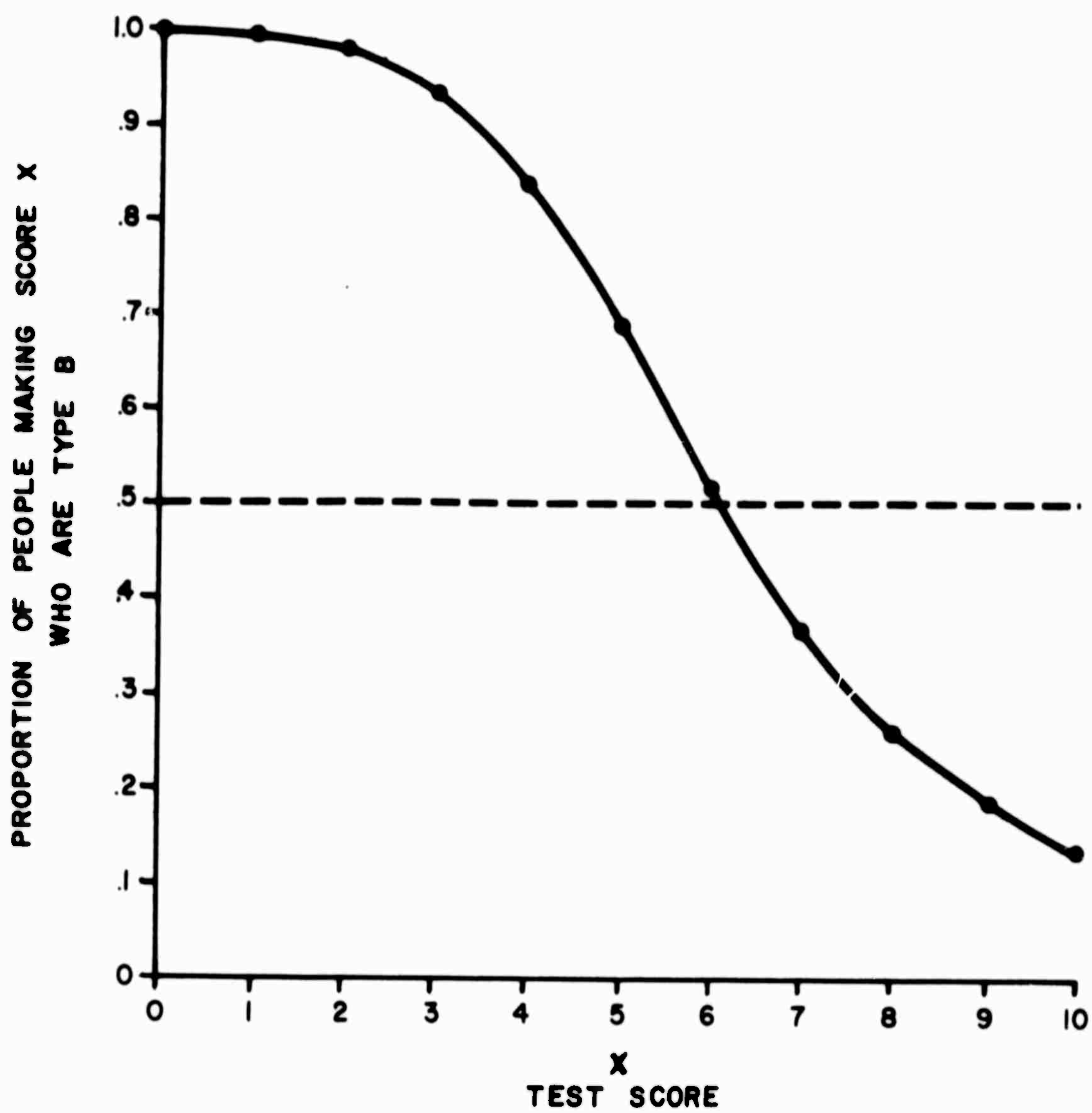

Figure 9. Discrimination unrelated to achievement or ability level. Half of the tested population are of Type A and guess with probability of chance success of ½ while the other half are of Type B and do not guess.

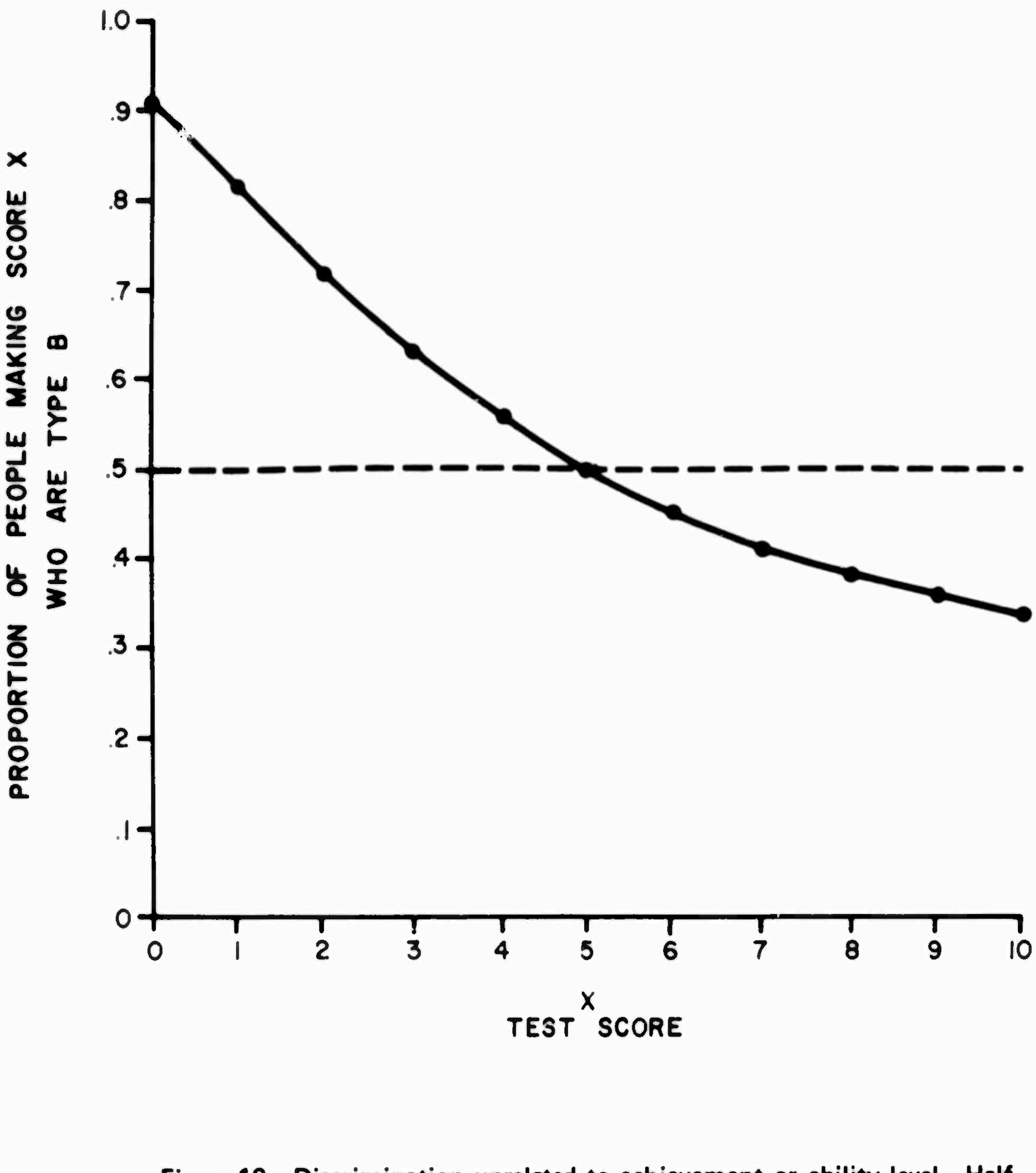

Figure 10. Discrimination unrelated to achievement or ability level. Half of the tested population are of Type A and guess with probability of chance success of 1/5 while the other half are of Type B and do not guess.

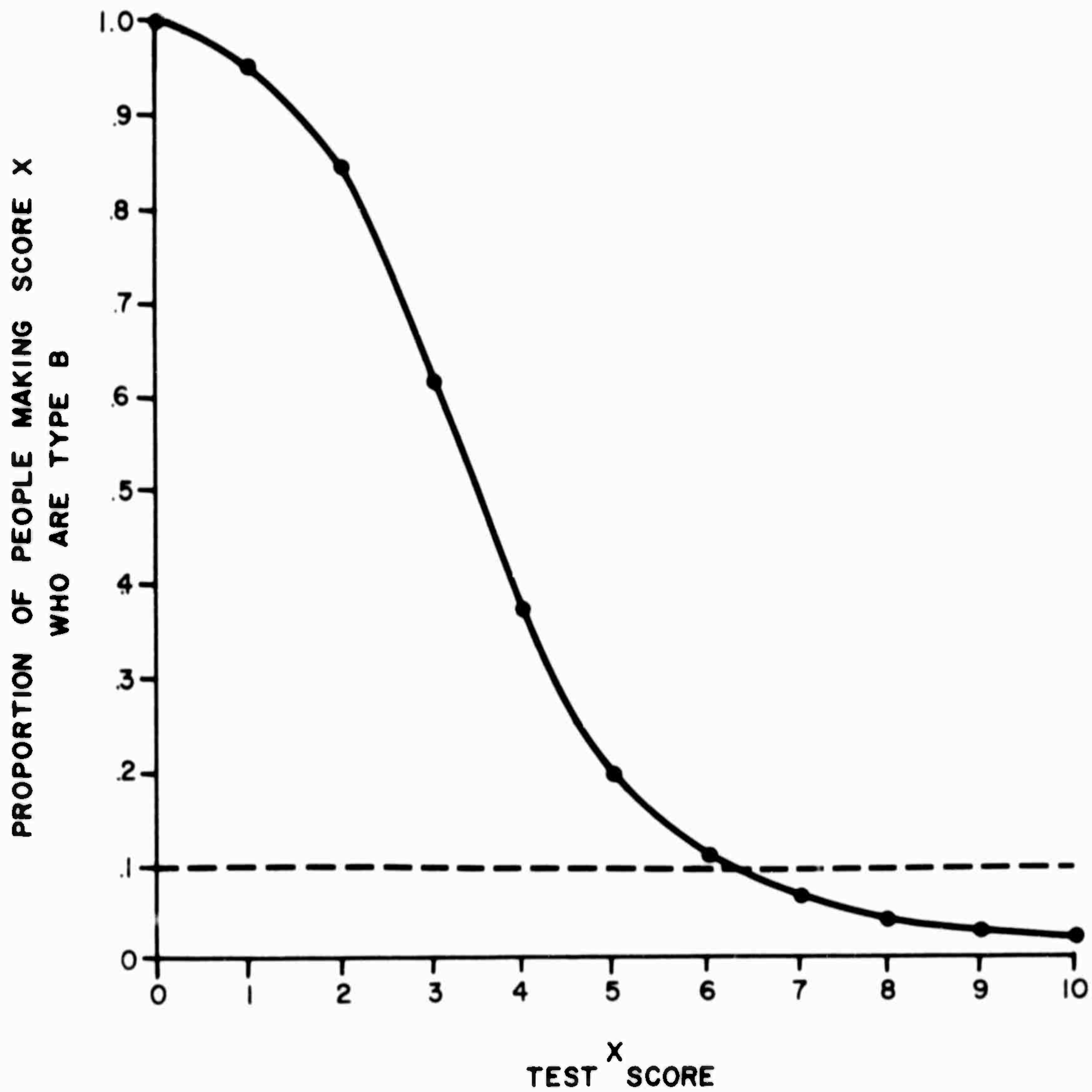

Figure 11. Discrimination unrelated to achievement or ability level. Ninety percent of the tested population are of Type A and guess with probability of chance success of ½ while the other ten percent are of Type B and do not guess.

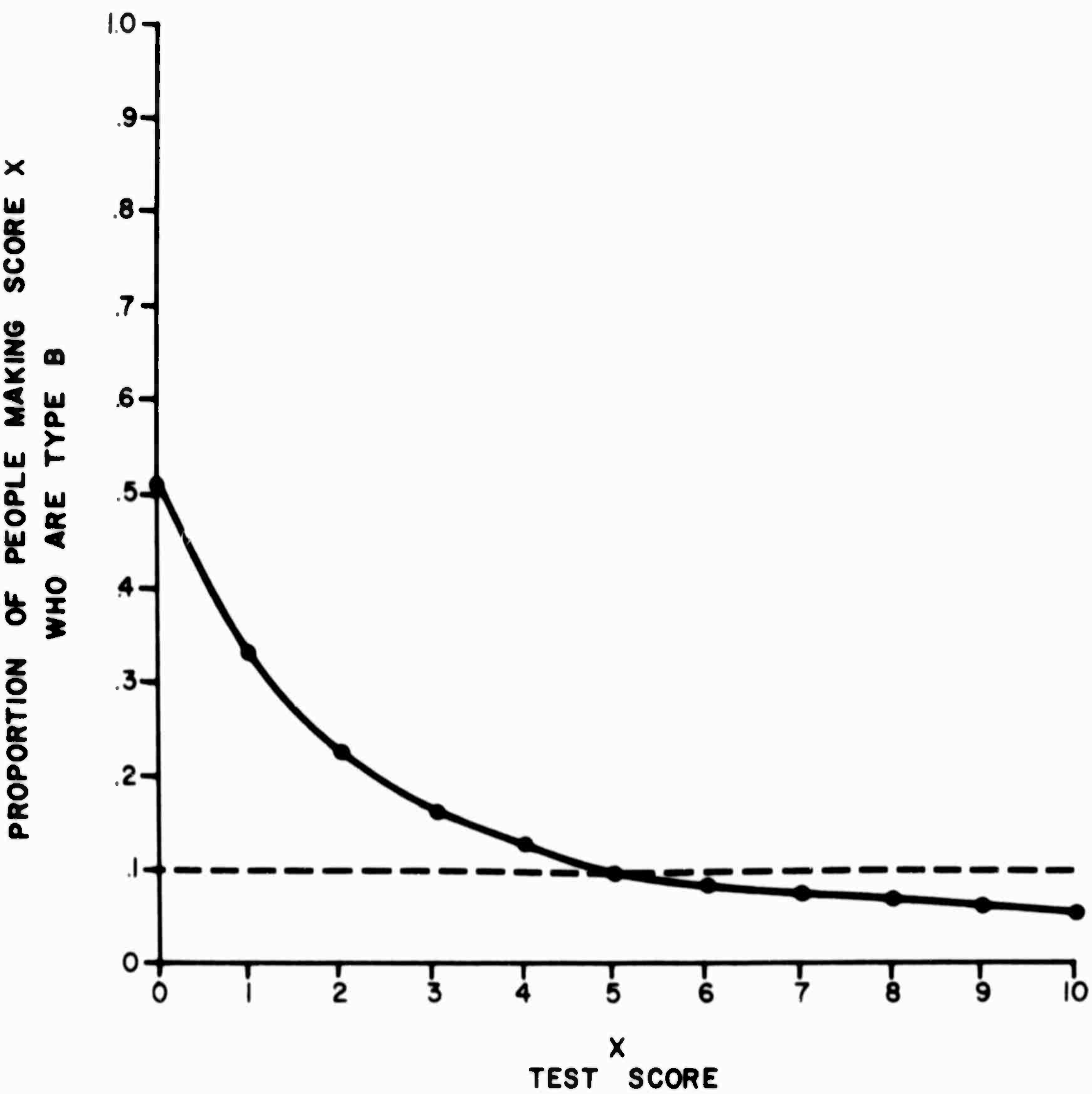

Figure 12. Discrimination unrelated to achievement or ability level. Ninety percent of the tested population are of Type A and guess with probability of chance success of 1/5 while the other ten percent are of Type B and do not guess.

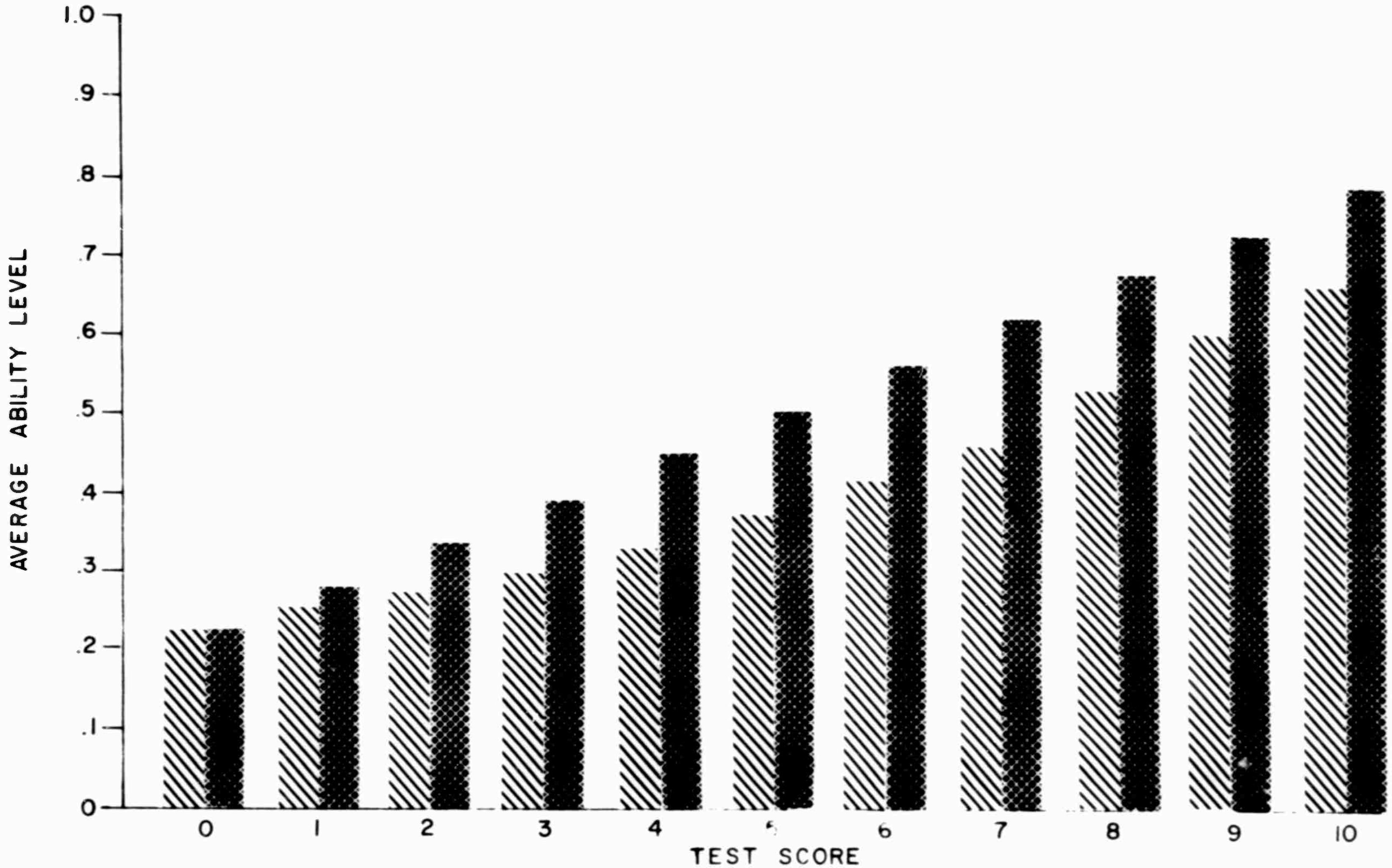

Figure 13. The superiority of Type B persons after placement. Left hand bars: Type A persons guessing with probability of chance success of ½. Right hand bars: Type B persons. For any test score other than zero, a Type B person will on the average, have more ability or greater achievement than will the Type A person making the same score.

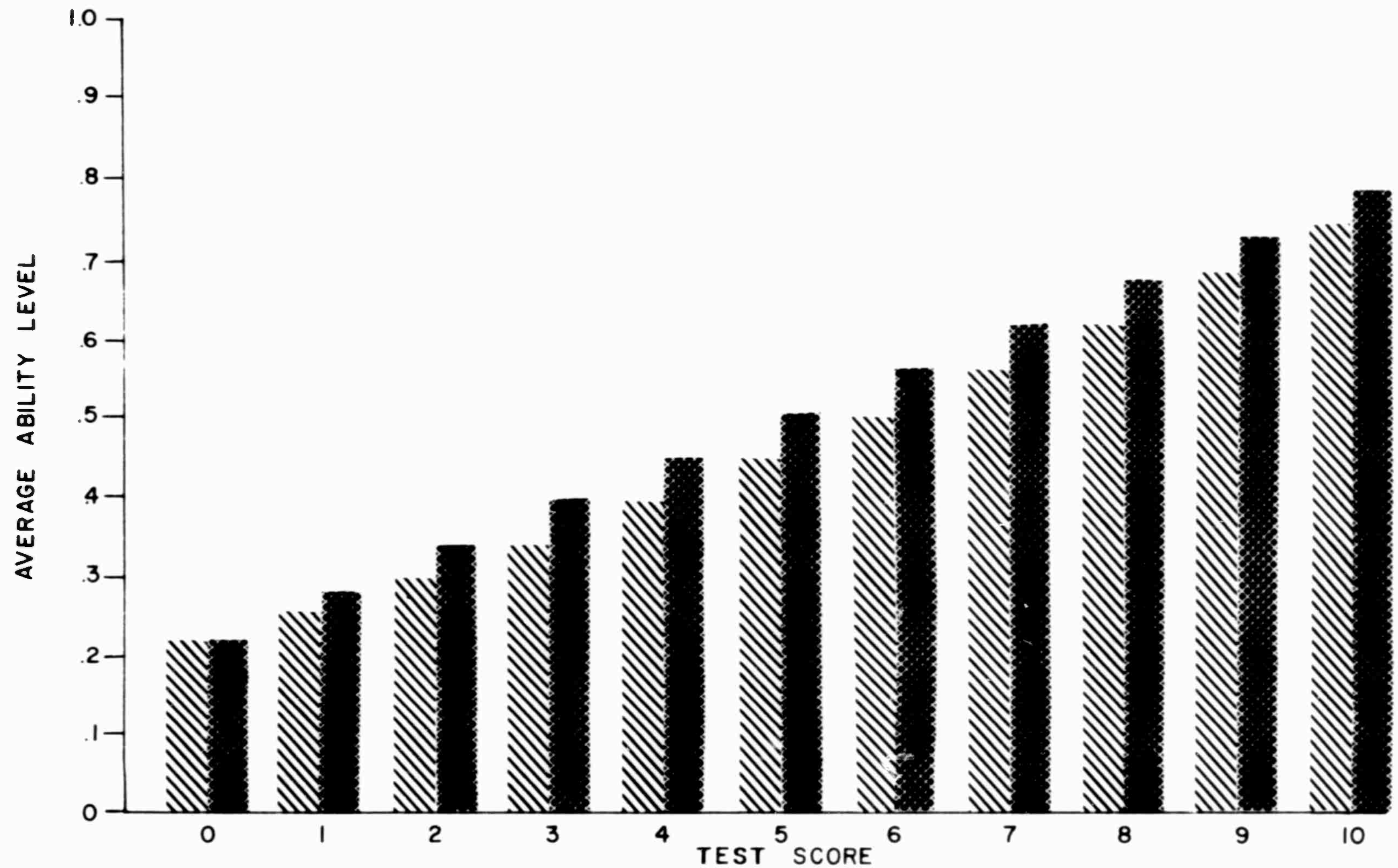

Figure 14. The superiority of Type B persons after placement. Left hand bars: Type A persons guessing with probability of chance success of 1/5. Right hand bars: Type B persons. For any test score other than zero, a Type B person will on the average, have more ability or greater achievement than will the Type A person making the same score.

Figure 15. Performance of a ten-item objective test in selecting people of above average ability when no one guesses. Upper bars: Fifty percent of tested population who have above average ability. Lower bars: Fifty percent of tested population who have below average ability.

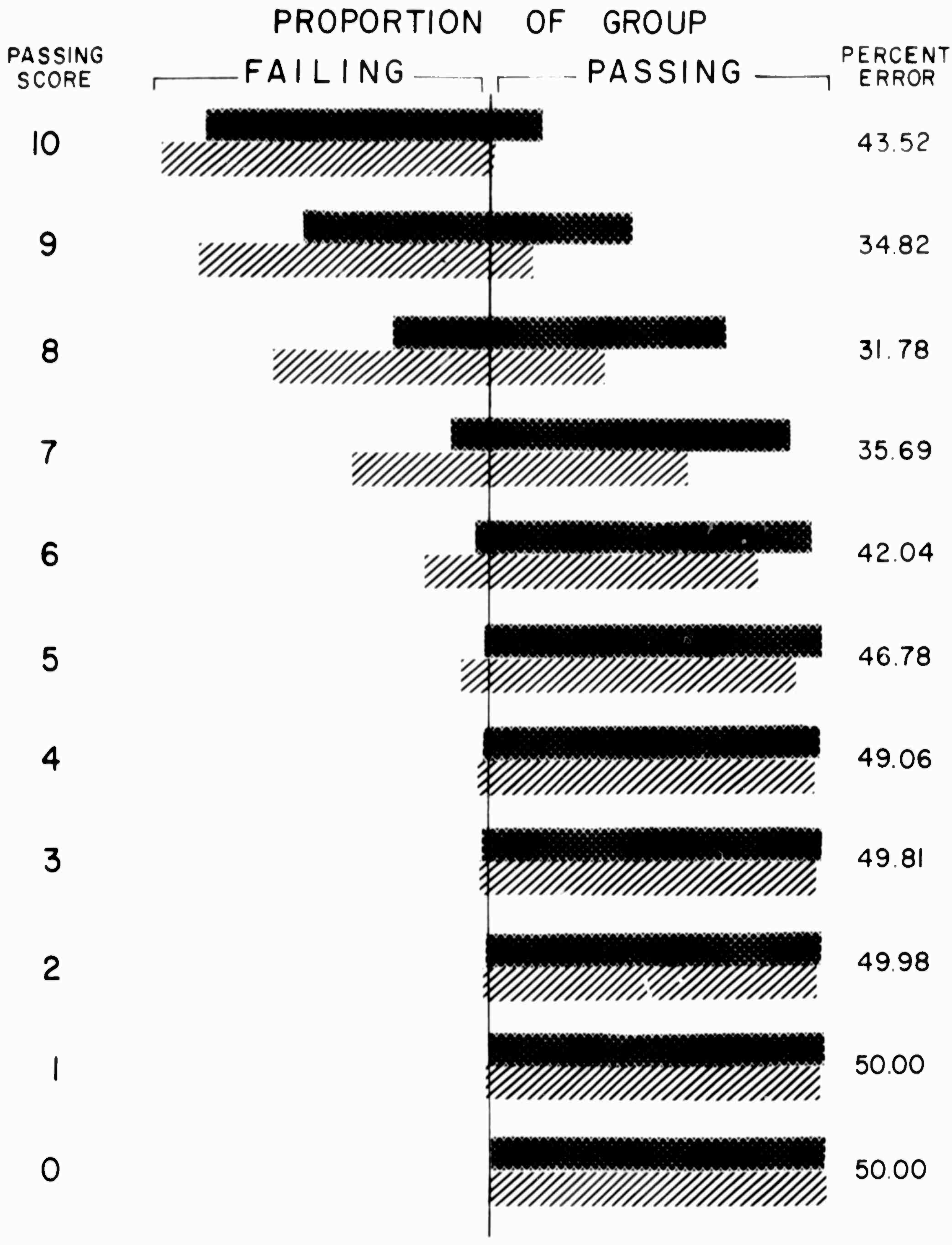

Figure 16. Performance of a ten-item objective test in selecting people when every one guesses with probability of chance success of ½. Upper bars: Fifty percent of tested population who have above average ability. Lower bars: Fifty percent of tested population who have below average ability.

Figure 17. Performance of a ten-item objective test in selecting people when half of the population guesses with probability of chance success of ½ while the other half do not guess. Upper bars: Fifty percent of tested population who have above average ability. Lower bars: Fifty percent of tested population who have below average ability.

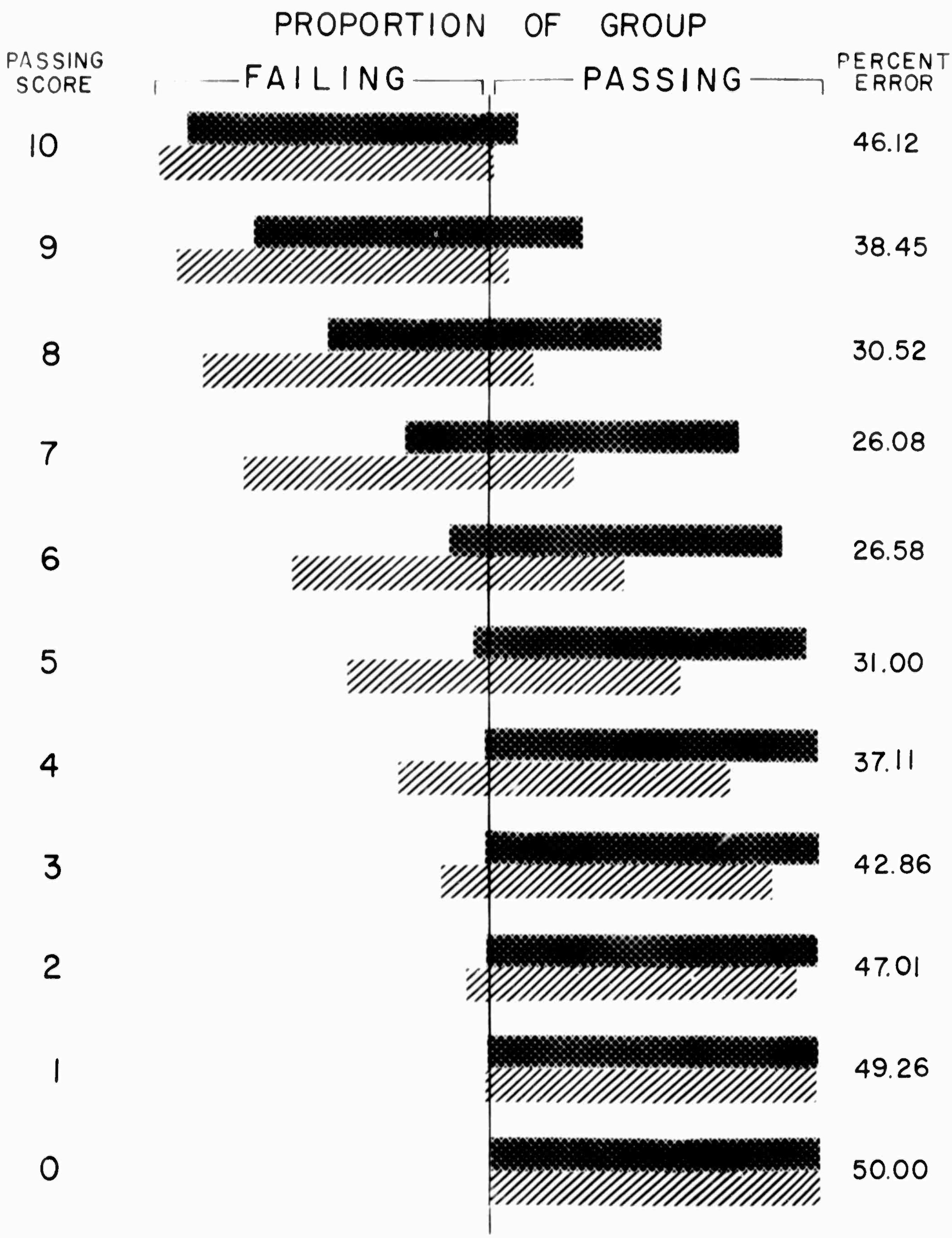

Figure 18. Performance of a ten-item objective test in discriminating people who are guessing from those who don't. Conditions exactly the same as for Figure 17. Upper bars: Half of tested population who guess with probability of chance success of ½. Lower bars: Half of tested population who do not guess.

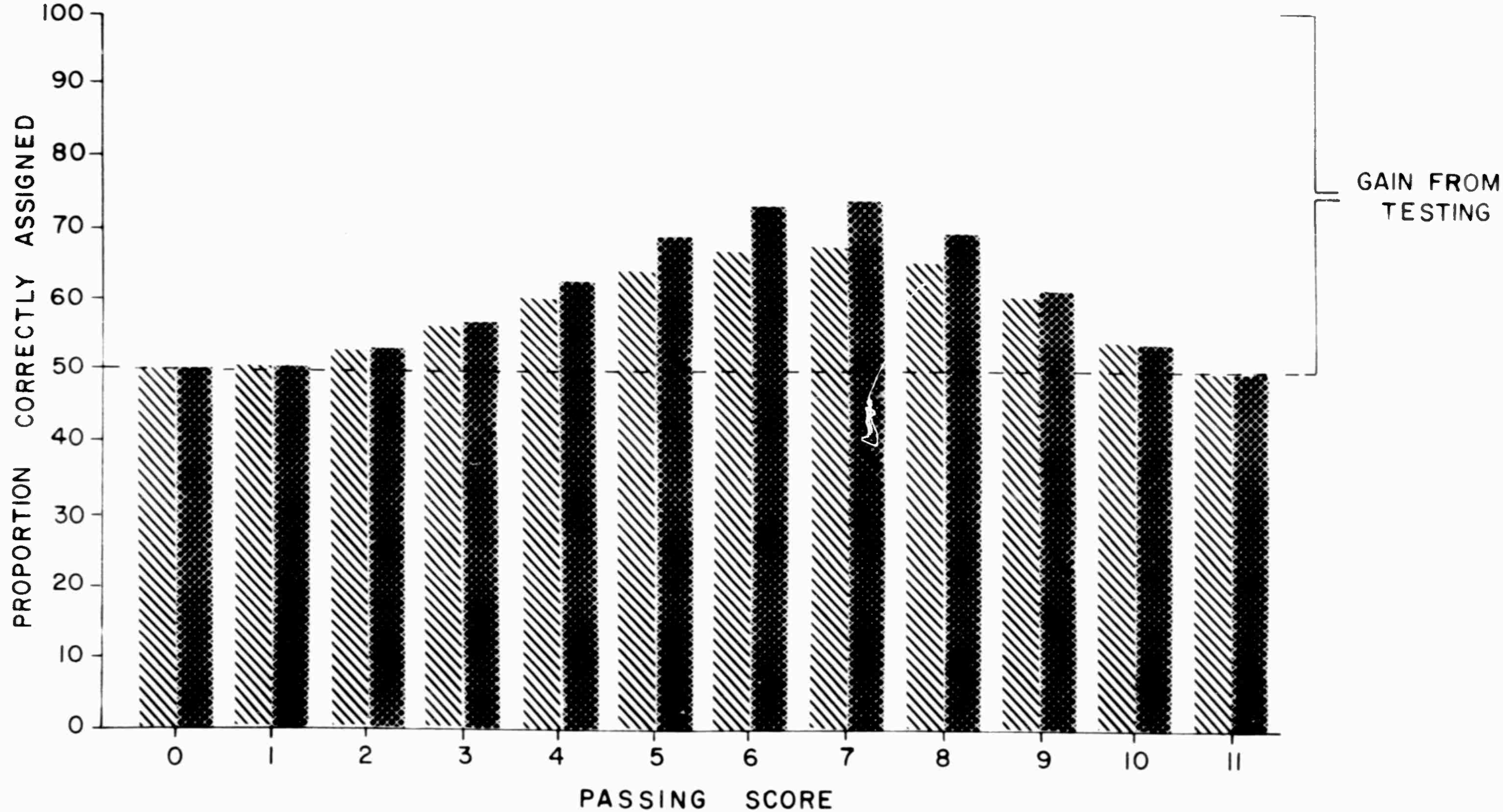

Figure 19. Selecting on the basis of ability or achievement vs. discriminating against those who do not guess. Conditions exactly the same as those for Figures 17 and 18. Left hand bars: Performance of ten-item objective test in selecting people of above average ability. Right hand bars: Performance of ten-item test in discriminating against those who do not guess.

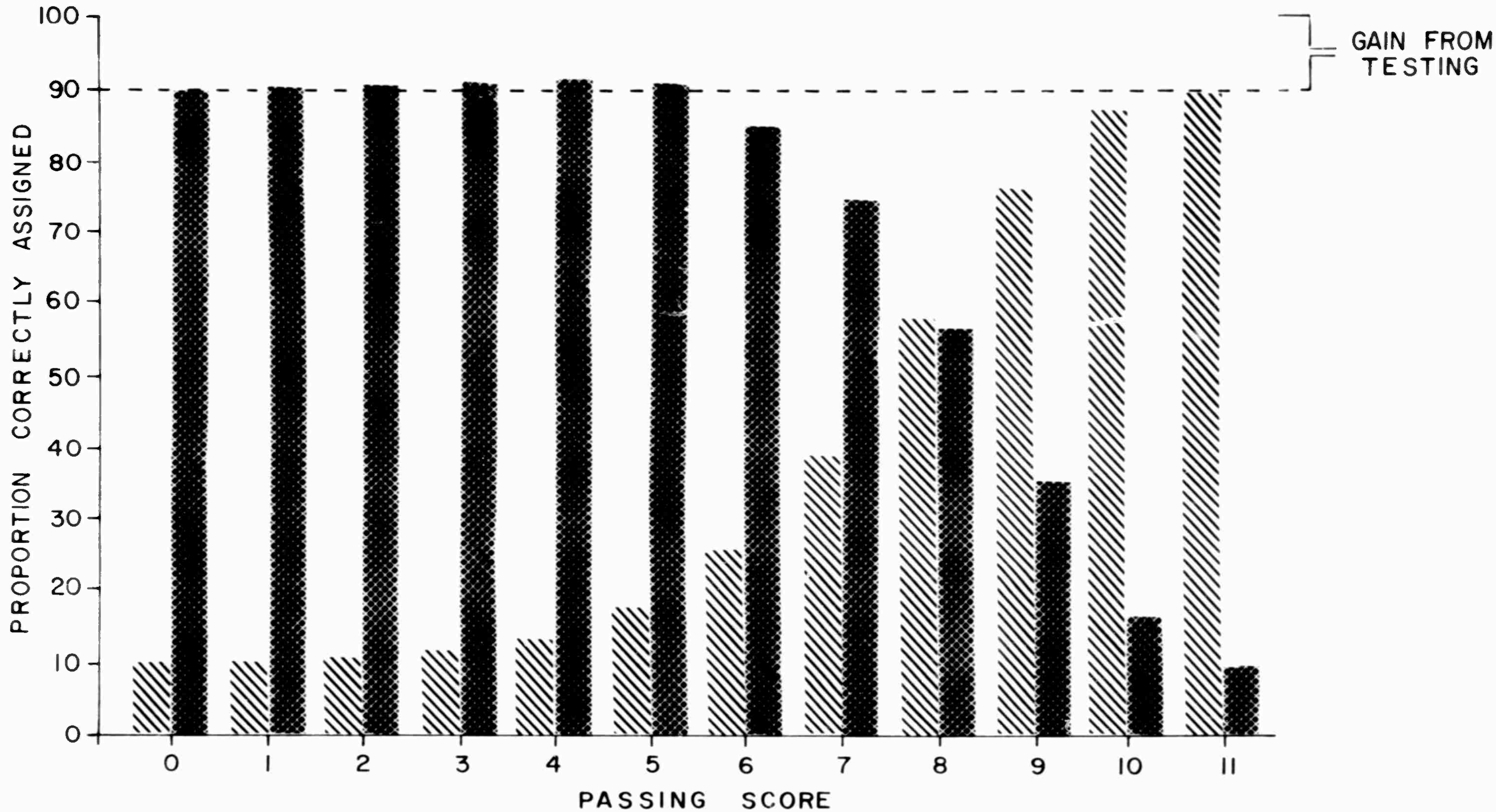

Figure 20. Selecting on the basis of ability or achievement vs. discriminating against those who do not guess. Ninety percent of the tested population guesses with probability of chance success of ½. Left hand bars: Performance of ten-item objective test in selecting people in upper ten percent of ability or achievement levels. Right hand bars: Performance of ten-item test in discriminating against those who do not guess.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The Shuford-Massengill Corpnration Box 26, Lexington, Massachusetts 02173	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

INDIVIDUAL AND SOCIAL JUSTICE IN OBJECTIVE TESTING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Emir H. Shuford, Jr. & H. Edward Massengill

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1967	30	8

8a. CONTRACT OR GRANT NO. AF 49(638)-1744 b. PROJECT NO. 920F - 9719 c. 6154501R d. 681313	9a. ORIGINATOR'S REPORT NUMBER(S) SMC R-10 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR 68-2160

10. DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research (SRLB) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

With the development of decision-theoretic psychometrics and Valid Confidence Testing, it is now possible to administer objective and semi-objective tests in such a way that guessing is practically eliminated from the test data. In order to estimate the benefits available from the use of these new procedures, it becomes important to estimate the effects of guessing upon test data obtained by using the old methods of administration.

Logic and mathematics are used to examine the effects of guessing upon an examinee's test score. The decision to guess or not to guess in taking an examination has a great effect on the examinee's chances of passing the test.

The possibility that the examinee's decision to guess on taking an objective examination may be a function of his previous educational experience and cultural background is considered, as a possible source of cultural bias. It is shown mathematically that when two groups of examinees with exactly the same distribution of ability levels take the same test and if the examinees in one group guess while the others don't, the guessing group will appear in the test results as having more ability than the examinees in the other group. It is shown further that of those examinees making the same test score, the examinees who do not guess will have a higher average ability level than those who do guess.

It is possible to evaluate the performance of an objective test not only according to its ability to discriminate between examinees of different ability levels but also according to its ability to discriminate between those who are guessing and those who are not guessing. It is shown that an objective test can be a better test of cultural and educational background than it is of ability level. Thus, the social justice of the older forms of objective test leave much to be desired. These injustices may be eliminated through the application of decision-theoretic psychometrics.

DD FORM 1473

UNCLASSIFIED

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Objective testing						
Guessing						
Cultural bias						
Decision-theoretic psychometrics						
Valid Confidence Testing						