DOCUMENT RESUME

ED 111 858 TM 004 825

AUTHOR Petersen, Nancy S.; Novick, Melvin R.
TITLE An Evaluation of Some Models for Culture-Fair Selection.
PUB DATE [Apr 75]
NOTE 56p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$3.32 Plus Postage
DESCRIPTORS *Comparative Analysis; *Culture Free Tests; *Models; *Personnel Selection; Prediction; Predictive Ability (Testing); Probability; Racial Discrimination; *Test Bias; Testing Problems

ABSTRACT

Models proposed by Cleary, Thorndike, Cole, Linn, Einhorn and Bass, Darlington, and Gross and Su for analyzing bias in the use of tests in a selection strategy are surveyed. Several additional models are also introduced. The purpose is to describe, compare, contrast, and evaluate these models while extracting such useful ideas as may be found in these approaches. The models of Thorndike, Cole, and Linn are judged to contain operational contradictions because of their use of the wrong conditional probability within the context of the probabilistic structure. These models, deriving from a concept of group parity, are also shown to have highly objectionable practical implications. It is suggested that the use of any of these models is contraindicated and that, indeed, the very concept of culture-fair selection is unworkable. It is then suggested that the necessary level of compensatory treatment for disadvantaged persons can be guaranteed only through the formal use of an appropriate model based on the Von Neumann-Morgenstern theory of maximizing expected utility. The models of Cleary, Einhorn and Bass, Gross and Su are based on what we judge to be the correct conditional probability and are special cases of the Expected Utility Model, but each has limited applicability. (Author/RC)

AN EVALUATION OF SOME MODELS FOR CULTURE-FAIR SELECTION

by

Nancy S. Petersen and Melvin R. Novick†
The University of Iowa

Abstract

In this paper, we survey models proposed by Cleary, Thorndike, Cole, Linn, Einhorn and Bass, Darlington, and Gross and Su for analyzing bias in the use of tests in a selection strategy. Several additional models are also introduced. Our purpose is to describe, compare, contrast and evaluate these models while, at the same time, extracting such useful ideas as may be found in these approaches. Several of these models (those of Thorndike, Cole and Linn) are judged to contain operational contradictions because of their use of the wrong conditional probability within the context of the probabilistic structure. These models, deriving from a concept of group parity, are also shown to have highly objectionable practical implications. It is suggested that the use of any of these models is contraindicated and that, indeed, the very concept of culture-fair selection is unworkable. It is then suggested that the necessary level of compensatory treatment for disadvantaged persons can be guaranteed

---

†This paper derives from material contained in a thesis submitted by Nancy S. Petersen in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Iowa, July, 1974.

only through the formal use of an appropriate model based on the Von Neumann-Morgenstern theory of maximizing expected utility. Three of the models studied (Cleary, Einhorn and Bass, Gross and Su) are based on what we judge to be the correct conditional probability and are special cases of the Expected Utility Model, but each has limited applicability.

Introduction

Tests are being used extensively by businesses and educational institutions for the screening of applicants for jobs or training programs. A major problem facing these institutions is how to eliminate cultural or racial unfairness arising from the use of tests in this process. There are many different definitions of what constitutes culture-fair selection, each implicitly, though unfortunately not explicitly, involving a particular set of value judgments with different implications for how selection should be accomplished. For each of these definitions a remedy has been proposed. Our purpose in this paper is to show that some of these approaches are inadequate to their task and that more complex analyses are required.

Description of the Selection Process

The selection process can be characterized in the same manner for all selection models. First, there are individuals about whom decisions are required. These decisions are to be based on information about the individuals. The information is processed by some strategy which leads to a final decision. The final decision ends the decision-making process by assigning the individual to either a selected or an outselected group. The outcome is the individual's performance after the assignment or, in other words, the consequences resulting from the decision. (Cronbach and Gleser, 1965, p. 18.)

A strategy is a rule for making decisions. Each selection model represents a strategy, the intent of which is to guarantee cultural

fairness in the selection process. Information is generally provided by a test and we shall use the term test to refer to all information-gathering procedures including interviews and physical measurements. The overriding problem is the lack of agreement as to the meaning of the term culture-fair selection.

Each of the selection models or strategies we discuss can be characterized in the same manner. It is assumed that the applicants to an educational institution, to a training program or for employment can be separated into subpopulations because of an a priori belief that the assumed linear regressions within these subpopulations are different:-that is, the test (or predictor) may be more valid for some subpopulations than for others (different slopes), and/or for a fixed value of the predictor, the level of criterion performances may differ (different intercepts), or that some differential selection criterion is appropriate for various subpopulations. Alternatively, these subpopulations may be differentiable primarily because of public concern with what is going on in them and a public need to verify that all subpopulations are being handled "fairly". Specifically, it is assumed that initially a criterion score (Y), as well as a predictor or test score (X), is available for members of each subpopulation, implying that in the past all applicants have been admitted or employed regardless of their score on the test. A minimum level of satisfactory criterion performance (y*) is determined. The number of applicants that can be selected is specified. If there is no constraint on the number of applicants that can be accepted, then the selection situation is referred to as quota-free selection; if

only a fixed proportion of the applicants can be accepted, then the selection situation is referred to as restricted selection. A cut score (x^*) on the predictor needs then to be calculated for each subpopulation so that the definition of culture-fair selection specified by the particular model is satisfied. In the case of multiple predictors or tests (X_1, X_2, ..., X_m), the cut score will be determined on the variable formed by the usual least squares linear combination of the predictor variables. In the future, applicants with a test score above the predictor cut score for their subpopulation will be selected and applicants with a test score below the predictor cut score for their subpopulation will be rejected.

This selection strategy presupposes that an acceptable criterion variable is available. The inappropriateness of the criterion variable will not be treated in this paper, although this may be the most important problem. Thus, the following discussion of selection fairness will be based on the premise that the available criterion score is a relevant, reliable and unbiased measure of performance for applicants in each subpopulation. All previous contributions in this area have implicitly made these assumptions. Caveat emptor!

The Regression Model

The Regression Model for "test bias" has been well stated by Cleary (1968):

> A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent non-zero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup. With this definition of bias, there may be a connotation of "unfair," particularly if the use of the test produces a prediction that is too low. [p. 115.]

Here, standard regression theory with the minimization of mean-squared error is used to provide predicted criterion scores. At the minimum level of satisfactory criterion performance (y*),

$$y^* = \alpha_1 + \beta_1 x_1^* = \ldots = \alpha_g + \beta_g x_g^* , \qquad (1)$$

where α_i, β_i, and x_i^* represent the intercept, slope and predictor cut score for subpopulation $\pi_i (i = 1, \ldots, g)$, respectively. If the regression lines are identical in each subpopulation, then the use of the common regression equation to select applicants with the highest predicted criterion scores is considered fair.

Using the Regression Model, and assuming that the parameters (α_1, β_1), (α_2, β_2), ..., (α_g, β_g) are known precisely, a decision maker can be assured that the average predicted criterion score, given the available predictor variables, will be a maximum for the applicants selected and, incidentally, a minimum for the applicants rejected. Using the Regression Model, the applicants can be assured that the selection procedure is "fair" to individual members of each subpopulation in that criterion performance is not systematically under or overpredicted for members of any subpopulation. Or to put it another way, the Regression Model says that if two applicants are being considered for one post, then that applicant having the highest predicted performance would be selected with prediction being made on the basis of subpopulation regression.

This definition of culture-fair selection assumes fairness is achieved if the applicants with the highest predicted criterion scores, using separate regression equations within subpopulations, are selected. From this point of view, selection is fair if and

only if it is based on the best prediction available. Thus, optimal prediction and fairness (lack of bias) are taken to be strictly equivalent and the procedure adopted is that which maximizes expected performance for each individual and hence overall.

To illustrate, suppose the applicants to an institution can be divided into two subpopulations referred to as subpopulation π_1 and subpopulation π_2. Now, refer to Figure 1. In Figure 1(a), the regression lines for the two subpopulations have the same slope but different intercepts. In Figure 1(b), the regression lines for the two subpopulations have different slopes and different intercepts with the point of intersection outside the range of possible test scores. In each of these situations, suppose the common regression line (π_c) for the total applicant population were used for predicting criterion scores for all applicants rather than the separate within-subpopulation regression lines, then for any given test score, criterion scores for subpopulation π_2 would be consistently underpredicted and, therefore, this subpopulation would be discriminated against by the test. In Figure 1(c), the regression lines for the two subpopulations again have different slopes and different intercepts, but the point of intersection is inside the range of possible test scores. If the common regression line were used for predicting criterion scores, then some individuals from both subpopulations would be discriminated against. At point x_1 on the test, the criterion score for a member of subpopulation π_1 would be underpredicted and, at point x_2 on the test, the criterion score for a member of subpopulation π_2 would be underpredicted. In the special case in which the regression lines for the two subpopulations coincide, an applicant's predicted criterion score is the

Figure 1

Illustration of Culture-fair Selection

as Defined by the Regression Model

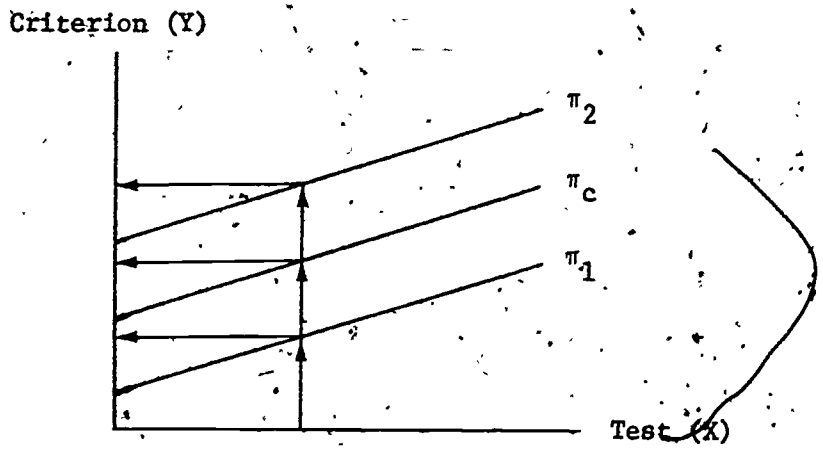


Figure 1(a). Supopulations with parallel regression lines but different intercepts.

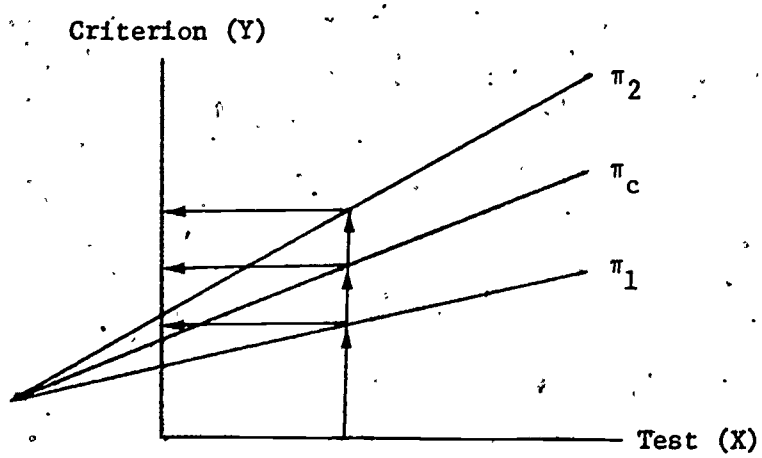


Figure 1(b). Subpopulations with different regression lines. Point of intersection outside range of possible test scores.

Figure 1 (cont'd.)

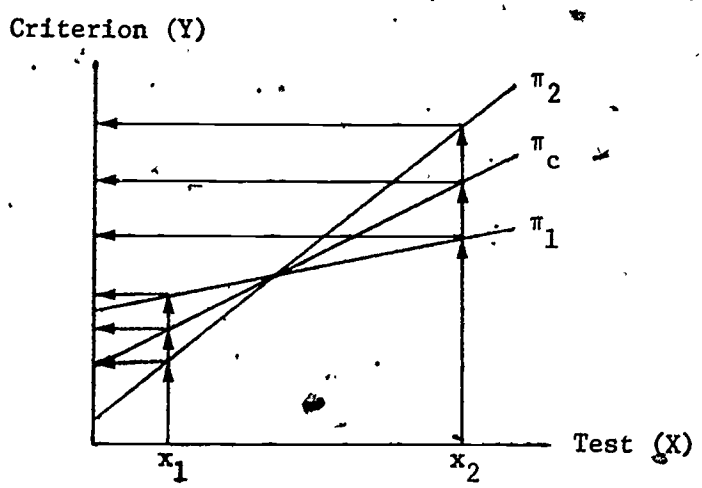


Figure 1(c). Subpopulations with different regression lines. Point of intersection inside range of possible test scores.

same, regardless of group membership, for any given test score.

The Regression Model is the most widely used model of selection fairness within the predictive context. It is a straightforward application of minimum-mean-squared error theory. It has been used in a number of empirical studies (e.g., Cleary, 1968; Bowers, 1970; Temp, 1971) and it has been basic in the conceptualizations and discussions of selection fairness that may be found in Anastasi (1968), Guion (1966), Bartlett and O'Leary (1969), Einhorn and Bass (1971), Linn and Werts (1971), Linn (1973), and Schmidt and Hunter (1974).

The Constant Ratio Model

Thorndike (1971) suggests that in a study of culture-fair selection we should consider the implications for the proportions of applicants admitted from each subpopulation as well as the implications of the within-subpopulation regression lines as was suggested by the Regression Model. He demonstrates that if a test has equal regression lines for each subpopulation but the discrepancy between subpopulations on the test differs from the discrepancy between subpopulations on the criterion, then the use of the selection strategy implied by the Regression Model,

> which is "fair" to individual members of the group scoring lower on the test, is "unfair" to the lower [scoring] group as a whole in the sense that the proportion qualified on the test will be smaller, relative to the higher-scoring group, than the proportion that will reach any specified level of criterion performance. [p. 63.]

Thorndike proposes that in a fair selection procedure,

> the qualifying scores on a test should be set at levels that will qualify applicants in the two groups in proportion to the fraction of the two groups reaching a specified level of criterion performance. [p. 63.]

This definition assumes that the selection procedure is fair if applicants are selected so that the ratio of the proportion selected to the proportion successful is the same in all subpopulations; hence the reference to it as the Constant Ratio Model. Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

$$R = \frac{\text{Prob}(X \geq x_1^* | \pi_1)}{\text{Prob}(Y \geq y^* | \pi_1)} = \ldots = \frac{\text{Prob}(X \geq x_g^* | \pi_g)}{\text{Prob}(Y \geq y^* | \pi_g)}, \quad (2)$$

where R is a fixed constant for all subpopulations π_i, and x_i^* represents the predictor cut score for subpopulation π_i ($i = 1, \ldots, g$). It should be noted that Thorndike did not give a formal statement of a model, only a general prescription. The explication of the model, as given above, is due to Cole (1973).

To illustrate, refer to Figure 2.† (Adapted from Thorndike, 1971, p. 66.) Assume the applicants to the institution were divided into two subpopulations, π_1 and π_2. Figure 2(a) depicts the situation which Thorndike refers to as being "fair" to individual members of the minority population π_1 but "unfair" to the minority

---

†To simplify the diagrams in Figure 2 and Figure 4, it is assumed that (1) the variables X and Y have a bivariate normal distribution in each subpopulation, (2) the correlation between X and Y (r_{xy}) is positive, and (3) the standard deviation of the test (s_x), the standard deviation of the criterion (s_y), and r_{xy} are constant for each subpopulation. Furthermore, the predictor cut score (x_2^*) for the majority population is chosen to be on the regression line (i.e., for subpopulation π_2, given $X = x_2^*$, then $Y = y^*$). The predictor cut score (x_1^*) for the minority population is then adjusted accordingly.

Figure 2

Illustration of Culture-fair Selection

as Defined by the Constant Ratio Model

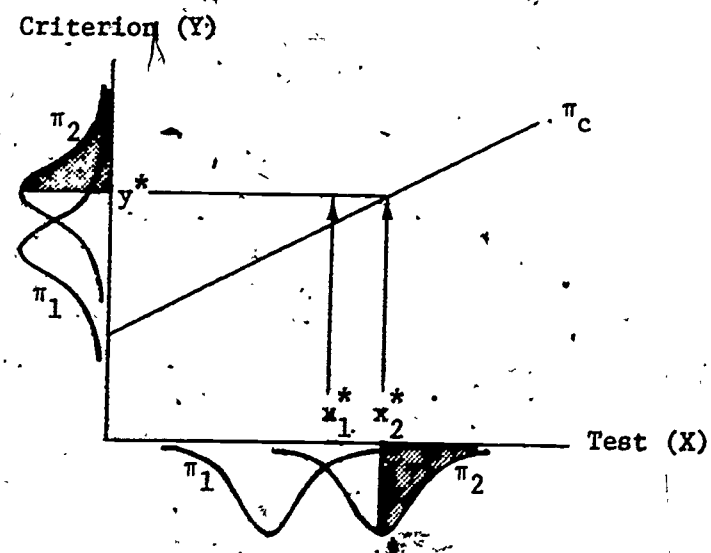


Figure 2(a). Subpopulations with common regression line. Mean difference on test is not equal to mean difference on criterion.

Figure 2 (cont'd.)

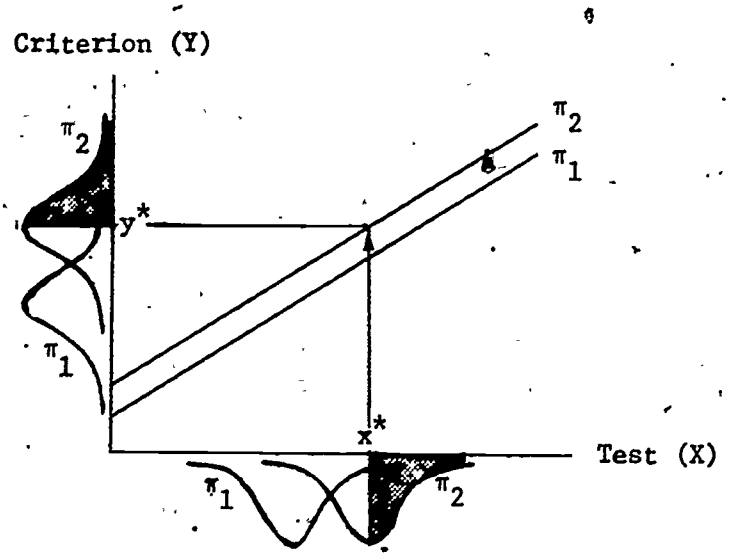


Figure 2(b). Subpopulations with parallel regression lines. Mean difference on test equals mean difference on criterion.

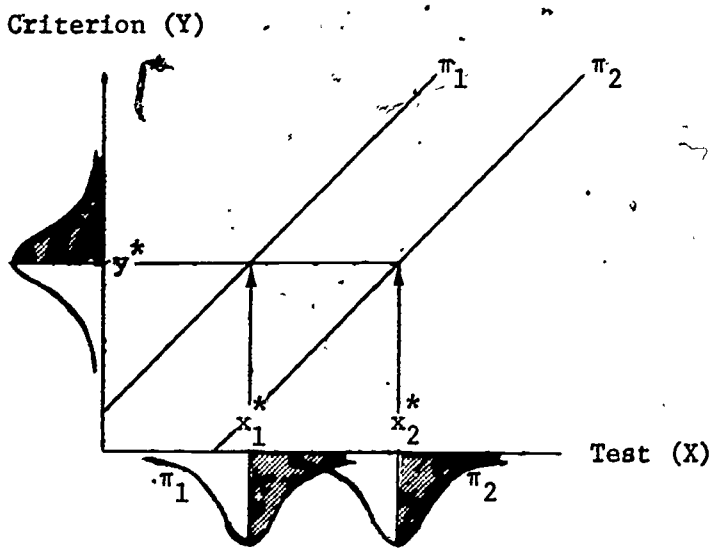


Figure 2(c). Subpopulations with parallel regression lines. Identical criterion score distributions.

population as a whole. The regression is identical in each subpopulation; thus the test would be considered fair according to the Regression Model if all individuals, regardless of group membership, who have test scores greater than or equal to x_2^* are selected. Note that the mean of X in subpopulation π_1 is less than in subpopulation π_2 and that this difference is greater than the corresponding difference on the criterion measure. If only those applicants with predicted criterion scores equal to or greater than y^* were selected, then approximately 50% of subpopulation π_2 would be accepted and approximately 50% would be successful, but essentially no members of subpopulation π_1 would be accepted, yet approximately 10% of the members of subpopulation π_1 would have been successful. Thus, if x_2^* is used as the predictor cut score for each subpopulation, the test discriminates against subpopulation π_1 according to the Constant Ratio Model. In this situation, to make the selection procedure fair according to the Constant Ratio Model, the members of subpopulation π_2 with test scores greater than or equal to x_2^* would be accepted, and members of subpopulation π_1 with test scores greater than or equal to x_1^* would be accepted. It is important to note, however, that if the difference in the group means on the predictor is less than that on the criterion measure, then application of the Constant Ratio Model will give the lower cut score to subpopulation π_2 rather than to π_1.

In figure 2(b), the regression lines are parallel and the difference between means on the test is the same as the difference between means on the criterion. The ratio of the proportion qualified on the test to the proportion successful is the same for each subpopulation. This strategy is fair according to the Constant

Ratio Model, but not according to the Regression Model. When there are differences in the distributions of the criterion scores, the strategy is fair according to both the Constant Ratio Model and the Regression Model only if the validity is perfect and the regression lines are the same for each subpopulation.

In Figure 2(c), the regression lines are parallel and the distribution of criterion scores is the same for both subpopulations. If y^* represents the minimum level of satisfactory criterion performance, then the same selection strategy would be considered fair by both the Regression Model and the Constant Ratio Model. The institution would accept members of subpopulation π_1 who had test scores greater than or equal to x_1^* and it would accept members of subpopulation π_2 who had test scores greater than or equal to x_2^*.

In many applications, the mean criterion score of the minority population π_1 will be less than in the majority population π_2 and that difference will be less than the difference of the predictor means, in which case an acceptance procedure based on the Constant Ratio Model will almost always accept applicants from the minority population π_1 who do less well on the criterion, on the average, than applicants from the majority population π_2. This feature explains the attractiveness of this model.

The Constant Ratio model and the Conditional Probability and Equal Probability models, to follow, have been described (Sawyer, Cole and Cole, 1975) as Group Parity models in that they focus on fairness to groups rather than to individuals. By contrast other models studied focus on the individual.

The Conditional Probability Model

Cole (1973) proposed a fully explicated criterion for culture-fair selection based on the conditional probability of being selected given satisfactory criterion performance; hence the reference to it as the Conditional Probability Model. Cole argues that all applicants who, if selected, are capable of being successful should be guaranteed an equal, or fair, opportunity to be selected, regardless of their group membership.

> The basic principle of the conditional probability selection model is that for both minority and majority groups whose members can achieve a satisfactory criterion score [$Y \geq y^*$] there should be the same probability of selection regardless of group membership. [p. 240.]

Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

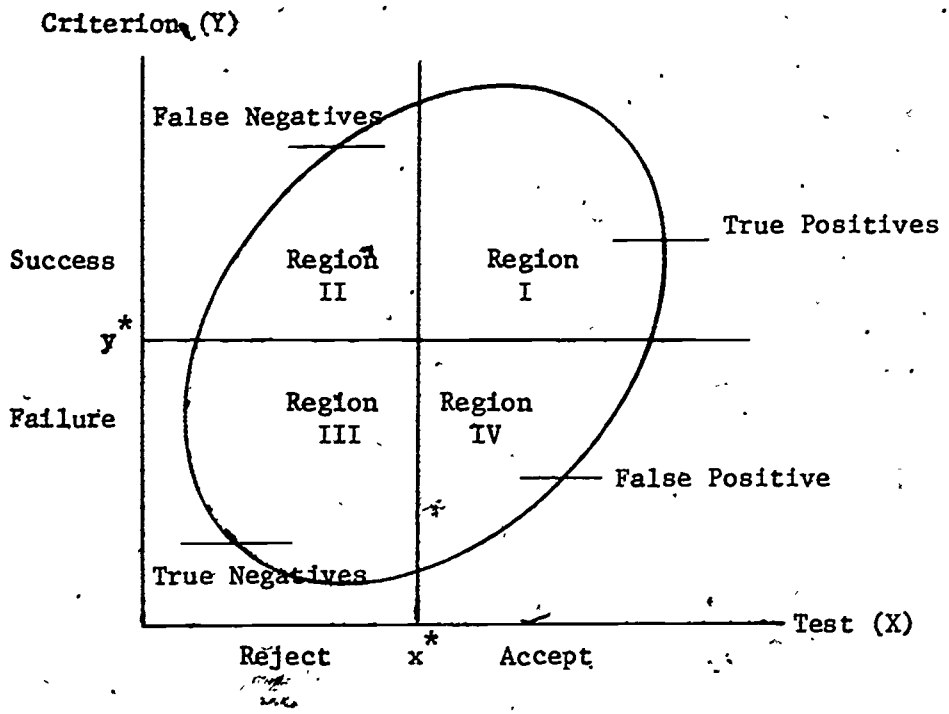
$$K = \text{Prob}(X \geq x_1^* | Y \geq y^*, \pi_1) = \ldots = \text{Prob}(X \geq x_g^* | Y \geq y^*, \pi_g), \tag{3}$$

where K is a fixed constant for all subpopulations π_i and x_i^* represents the predictor cut score for subpopulation $\pi_i (i = 1, \ldots, g)$.

Figure 3 is an illustration of a hypothetical bivariate distribution of test and criterion scores. Individuals falling in region II have test scores less than the predictor cut score (they would be rejected); yet, if selected, they would have satisfactory criterion performance. Such individuals are referred to as false negatives. False positives are those individuals with test scores greater than the predictor cut score (they would be accepted) but with unsatisfactory criterion performance. Such individuals fall in

Figure 3

A Hypothetical Bivariate Distribution



region IV. The assignment of an individual to either region II or IV is an incorrect decision (error). Correct decisions are made for those individuals assigned to regions I and III. (Linn, 1973, pp. 152-153.)

The emphasis in the Conditional Probability Model is on the number of applicants in region I in relation to the number of applicants in regions I and II combined, whereas the emphasis in the Constant Ratio Model is on the number of applicants in regions I and IV combined in relation to the number of applicants in regions I and II combined.†

Figure 4 contrasts the Regression Model, the Constant Ratio Model and the Conditional Probability Model for the situation in which the regression is identical for each subpopulation, but the mean test score and the mean criterion performance is less for members of subpopulation π_1 than for members of subpopulation π_2. (See comment in reference to Figure 2.) In Figure 4(a), all applicants, regardless of group membership, who have test scores greater than x^*, are accepted. Using this selection strategy, the selection procedure would be considered fair according to

---

†Linn (1973, p. 153) stated that a test was fair according to Thorndike's definition of selection fairness (the Constant Ratio Model) if the number of individuals in region II equals the number of individuals in region IV. (See Figure 3.) Strictly speaking, the Constant Ratio Model does not require equality of region II and IV. However, the model will be satisfied and equality of regions II and IV will occur if and only if the selection-success ratio R [Equation (2)] equals 1 implying $\text{Prob}(X \geq x_i^* | \pi_i) = \text{Prob}(Y \geq y^* | \pi_i)$ for each subpopulation π_i. For purposes of heuristic comparison among models, we shall assume that this assumption holds.

Figure 4

A Constrast of the Regression, the Constant Ratio, and the Conditional Probability Models

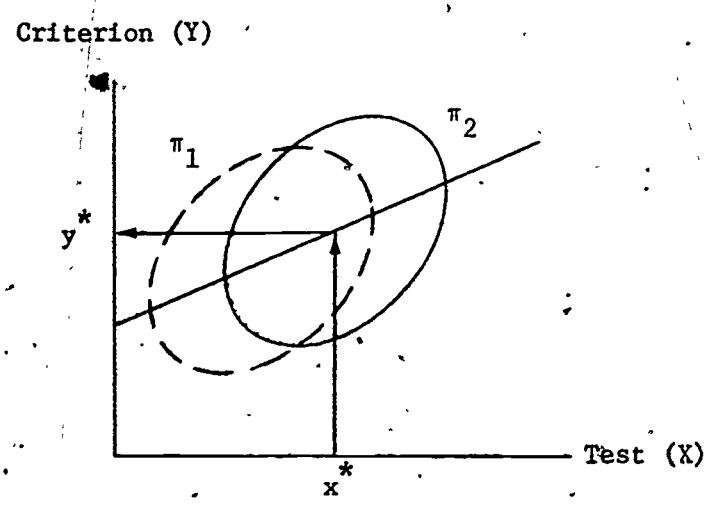


Figure 4(a). Subpopulations with common regression line. Selection strategy fair according to Regression Model.

Figure 4 (cont'd.)

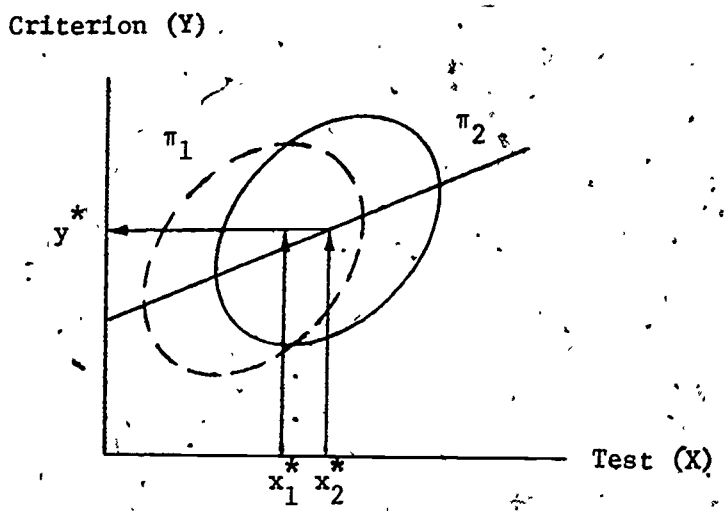


Figure 4(b). Subpopulations with common regression line. Selection strategy fair according to Constant Ratio Model.

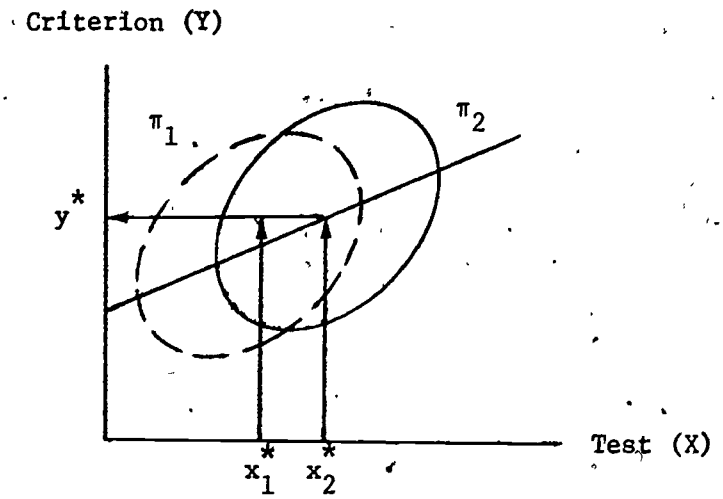


Figure 4(c). Subpopulations with common regression line. Selection strategy fair according to Conditional Probability Model.

the Regression Model. In Figure 4(b), applicants from subpopulation π_1 are accepted if they have test scores greater than x_1^*, and applicants from subpopulation π_2 are accepted if they have test scores greater than x_2^*. The ratio (I + IV)/(I + II) is constant for each subpopulation. (Refer to Figure 3.) Thus, using this selection strategy, the test is considered fair according to the Constant Ratio Model. In Figure 4(c), the predictor cut score for subpopulation π_1 is x_1^*, and for subpopulation π_2, the predictor cut score is x_2^*. Here the ratio I/(I + II) is constant for each subpopulation (refer to Figure 3). Using this selection strategy, the test is considered fair according to the Conditional Probability Model.

Note that as with the Constant Ratio Model, a selection strategy based on the Conditional Probability Model will almost always accept applicants from subpopulation π_1 who do less well on the criterion, on the average, than applicants from subpopulation π_2. Also note that if an applicant from subpopulation π_1 is predicted to do just as well on the criterion as an applicant from subpopulation π_2, then a selection strategy which is fair according to the Regression Model will consider the two applicants equally desirable candidates for admission. However, a selection strategy which is fair according to the Constant Ratio Model will give preference to the applicant from subpopulation π_1 and a selection strategy which is fair according to the Conditional Probability Model will given even greater preference to that applicant. Note also that this preference for subpopulation π_1 will hold even if the "minority" population happens to be majority subpopulation π_2, as will be discussed later. The remarks in this paragraph can be substantiated by noting that the

Cole criterion can be written as

$$\text{Prob}(X \geq x^* | Y \geq y^*) = \left[\frac{\text{Prob}(X \geq x^*)}{\text{Prob}(Y \geq y^*)}\right] \left[\text{Prob}(Y \geq y^* | X \geq x^*)\right] .$$

The first factor on the right is the Thorndike ratio and the second factor will be smaller for subpopulation π_1 than for π_2 under the specific conditions assumed earlier. Thus, for the Cole definition of selection fairness to be satisfied the cut score x^* for subpopulation π_1 must be lower than that value required to satisfy the Thorndike definition.

The Equal Probability Model

In the usual selection situation, the "given" information for each applicant is not his future state of being (success or failure) in relation to the criterion variable, but rather his present observed standing on the predictor variable. Thus, from one point of view, it would seem reasonable to propose a definition of culture-fair selection based on the conditional probability of success given selection. One might argue that all applicants who are selected should be guaranteed an equal, or fair, chance of being successful, regardless of group membership. Such a model for selection was described by Linn (1973, p. 153) and shall now be referred to as the Equal Probability Model.†

---

†Linn (1973, p. 153) described the Equal Probability Model but referenced it as the traditional psychometric approach suggested by Einhorn and Bass (1971). The definition of culture-fair selection suggested by Einhorn and Bass, to be called the Equal Risk Model, will be discussed later in the paper. At this point, it is enough

According to the Equal Probability Model, within the selected group from each subpopulation π_i, the proportion of successful performers should be the same. Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

$$Q = \text{Prob}(Y \geq y^* | X \geq x_1^*, \pi_1) = \ldots = \text{Prob}(Y \geq y^* | X \geq x_g^*, \pi_g) , \quad (4)$$

where Q is a fixed constant for all subpopulations π_i and x_i^* represents the predictor cut score for subpopulation π_i ($i = 1, \ldots, g$).

In reference to Figure 3, the emphasis in the Equal Probability Model is on the number of applicants in region I in relation to the number of applicants in regions I and IV combined. In reference to Figure 4, the selection strategy depicted in Figure 4(a) is fair, according to the Equal Probability Model, when members of subpopulation π_1 and members of subpopulation π_2 who are predicted to do equally well on the criterion are considered equally desirable candidates for admission. Clearly, a selection strategy dictated by the Equal Probability Model will not typically coincide with one derived from either of the three preceding models. Thus, the practitioner is faced with the task of choosing from among four "attractive" models. How should this choice be made?

---

to note that in the Equal Probability Model the conditioning is on $X \geq x_i^*$ while in the Equal Risk Model the conditioning is on $X = x_i^*$. It should be emphasized that the Equal Probability Model was not proposed by Linn, but was only discussed by him as an academic exercise. By contrast both the Constant Ratio and Conditional Probability models have been recommended for current adoption by Cole.

The Converse Constant Ratio Model

The last three models described (the Constant Ratio Model, the Conditional Probability Model and the Equal Probability Model) presented definitions of culture-fair selection stated in terms of success and selection probabilities. Conceptually, it seems just as reasonable to explicate the fundamental concept of each approach by exhibiting concern for the rejected and/or unsuccessful applicant. Thus, the following three models for culture-fair selection will be restatements of the previous three models in terms of failure and rejection.

Recall that the Constant Ratio Model compares selection rate with success rate in each subpopulation. The emphasis is on the proportion of applicants who are selected in relation to the proportion of applicants who are successful. However, one could conceivably consider it just as important or necessary to consider the implications for the proportion of applicants rejected in each subpopulation. One could propose that the cut scores on a test should be set at levels that will reject applicants in each subpopulation in proportion to the fraction of each subpopulation failing to reach a specified minimum level of criterion performance. Such a selection strategy will be referred to as the Converse Constant Ratio Model. If we are particularly concerned with a subpopulation π_i, then we would want to be sure that the ratio of the proportion rejected to the proportion failing be <u>no more</u> than in any other subpopulation, whereas with the original model we would want the ratio of the proportion selected to the proportion successful to be <u>not less</u> than in any other subpopulation.

The new definition assumes that a selection procedure is fair if applicants are rejected so that the proportion rejected to the proportion unsuccessful is the same in all subpopulations. Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

$$\bar{R} = \frac{\text{Prob}(X < x_1^* | \pi_1)}{\text{Prob}(Y < y^* | \pi_1)} = \ldots = \frac{\text{Prob}(X < x_g^* | \pi_g)}{\text{Prob}(Y < y^* | \pi_g)}, \quad (5)$$

where $\bar{R}$ is a fixed constant for all subpopulations π_i and x_i^* represents the predictor cut score for subpopulation π_i ($i = 1, \ldots, g$).

The above relationship can be rewritten as

$$\bar{R} = \frac{1 - \text{Prob}(X \geq x_i^* | \pi_i)}{1 - \text{Prob}(Y \geq y^* | \pi_i)}$$

$$= \frac{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - \text{Prob}(X \geq x_i^* | \pi_i)[\text{Prob}(Y \geq y^* | \pi_i)]^{-1}}{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1}$$

$$= \frac{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - R}{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1},$$

where $R = \text{Prob}(X \geq x_i^* | \pi_i)\ [\text{Prob}(Y \geq y^* | \pi_i)]^{-1}$ is the value to be equated among subpopulations for selection fairness as specified by the Constant Ratio Model. Now, suppose we have specified a minimum level of satisfactory criterion performance (y^*) and a selection-success ratio (R), then a predictor cut score x_i^* can be determined for each subpopulation π_i ($i = 1, \ldots, g$). Given the values y^*, R, and x_i^*, the rejection-failure ratio ($\bar{R}$) will be constant for each subpopulation

π_i if the following condition is satisfied:

$$\frac{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - R}{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1} = \frac{[\text{Prob}(Y \geq y^* | \pi_j)]^{-1} - R}{[\text{Prob}(Y \geq y^* | \pi_j)]^{-1} - 1}$$

for $i, j = 1, \ldots, g$.

The above condition will be satisfied if and only if either (1) $R = 1$ implying $\text{Prob}(X \geq x_i^* | \pi_i) = \text{Prob}(Y \geq y^* | \pi_i)$ for $i = 1, \ldots, g$, <u>or</u> (2) $\text{Prob}(Y \geq y^* | \pi_i) = \text{Prob}(Y \geq y^* | \pi_j)$ for $i, j = 1, \ldots, g$, but not generally. If either case (1) or case (2) obtains, the same set of predictor cut scores x_i^* ($i = 1, \ldots, g$) is considered fair according to both the Constant Ratio Model and its converse, but, otherwise, the strategies will differ. In Thorndike's illustration (1971, p. 66), he set $R = 1$, though he did not indicate that this was required by his model. Only by reference to various real applications might we be convinced that $R = 1$ will be a commonly acceptable value. However, with restricted selection, it is not generally possible to simultaneously satisfy this condition and the selection constraint.

Consider carefully the force of the following argument. If fairness to subpopulation π_i demands that the selection-success ratio (R) be the same for any other subpopulation π_j, then, <u>with identical logic</u>, fairness to subpopulation π_i demands that the rejection-failure ratio ($\overline{R}$) be the same for any other subpopulation π_j, and the two specifications cannot both be satisfied. This strikes us as being a logical contradiction; however, each reader must make a personal judgment on this point.

The Converse Conditional Probability Model

The Conditional Probability Model is based on the conditional probability of being selected given satisfactory criterion performance. The emphasis is on the proportion of potentially successful applicants who are selected. However, one could argue instead that potential failures should be rejected in no greater percentage in any subpopulation. We shall label this selection strategy the Converse Conditional Probability Model.

The Converse Conditional Probability Model is based on the conditional probability of being rejected given unsatisfactory criterion performance. Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

$$\overline{K} = \text{Prob}(X < x_1^* | Y < y^*, \pi_1) = \ldots = \text{Prob}(X < x_g^* | Y < y^*, \pi_g), \tag{6}$$

where $\overline{K}$ is a fixed constant for all subpopulations π_1 and x_1^* represents the predictor cut score for subpopulation π_i ($i = 1, \ldots, g$).

The above relationship can be rewritten as

$$\overline{K} = \text{Prob}(X < x_i^*, Y < y^* | \pi_i)[\text{Prob}(Y < y^* | \pi_i)]^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)] - \text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)\}$$

$$[1 - \text{Prob}(Y \geq y^* | \pi_i)]^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)][\text{Prob}(Y \geq y^* | \pi_i)]^{-1}$$

$$- \text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)[\text{Prob}(Y \geq y^* | \pi_i)]^{-1}\}\{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1\}^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)][\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - K\}$$

$$\{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1\}^{-1} ,$$

where $K = \text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)[\text{Prob}(Y \geq y^* | \pi_i)]^{-1}$ is the value to be equated among subpopulations for selection fairness as specified by the Conditional Probability Model. Now, suppose the decision maker has specified a minimum level of satisfactory criterion performance (y^*) and a constant conditional probability of selection given success (K), then a predictor cut score x_i^* can be determined for each subpopulation π_i ($i = 1, \ldots, g$). Given the values y^*, K, and x_i^*, the conditional probability of rejection given failure ($\bar{K}$) will be constant for each subpopulation π_i if the following condition is satisfied:

$$\{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)]$$

$$[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - K\} \{[\text{Prob}(Y \geq y^* | \pi_i)]^{-1} - 1\}^{-1}$$

$$= \{[\text{Prob}(X \geq x_j^*, Y \geq y^* | \pi_j) + \text{Prob}(X < x_j^*, Y < y^* | \pi_j)]$$

$$[\text{Prob}(Y \geq y^* | \pi_j)]^{-1} - K\}\{[\text{Prob}(Y \geq y^* | \pi_j)]^{-1} - 1\}^{-1}$$

for $i, j = 1, \ldots, g$.

This condition will be satisfied if and only if $\text{Prob}(Y \geq y^* | \pi_i) = \text{Prob}(Y \geq y^* | \pi_j)$ and $\text{Prob}(X < x_i^*, Y < y^* | \pi_i) = \text{Prob}(X < x_j^*, Y < y^* | \pi_j)$ for $i, j = 1, \ldots, g$, a most unlikely state of affairs. In that case, the same selection strategy, the same set predictor cut scores x_i^* $(i = 1, \ldots, g)$, is considered fair according to both the Conditional Probability Model and the Converse Conditional Probability Model, but, otherwise, the selection strategies will differ. It does not seem at all apparent to us whether for the particular subpopulation that is of public concern it is more important to keep the conditional probability of rejection (given potential failure) low or the conditional probability of acceptance (given potential success) high. Unfortunately, the two criteria are based on different conditioning events and hence are contradictory.

The Converse Equal Probability Model

The Equal Probability Model is based on the conditional probability of success given selection. The emphasis is on the proportion of the selected applicants who are successful. However, one could propose that all applicants who are rejected should have the same probability of being a failure, regardless of group membership. Such a selection strategy will be labeled as the Converse Equal Probability Model.

The Converse Equal Probability Model is based on the conditional probability of failure given rejection. Therefore, given a minimum level of satisfactory criterion performance (y^*), a selection procedure is considered fair when

$$\overline{Q} = \text{Prob}(Y < y^* | X < x_1^*, \pi_1) = \ldots = \text{Prob}(Y < y^* | X < x_g^*, \pi_g) , \tag{7}$$

where $\overline{Q}$ is a fixed constant for all subpopulations π_i and x_i^* represents the predictor cut score for subpopulation $\pi_i (i = 1, \ldots, g)$.

The above relationship can be rewritten as

$$\overline{Q} = \text{Prob}(X < x_i^*, Y < y^* | \pi_i)[\text{Prob}(X < x_i^* | \pi_i)]^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)] - \text{Prob}(X \geq x_i^*, Y > y^* | \pi_i)\}$$

$$[1 - \text{Prob}(X \geq x_i^* | \pi_i)]^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)][\text{Prob}(X \geq x_i^* | \pi_i)]^{-1}$$

$$-\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)[\text{Prob}(X \geq x_i^* | \pi_i)]^{-1}\}\{[\text{Prob}(X \geq x_i^* | \pi_i)]^{-1} - 1\}^{-1}$$

$$= \{[\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i) + \text{Prob}(X < x_i^*, Y < y^* | \pi_i)][\text{Prob}(X \geq x_i^* | \pi_i)]^{-1} - Q\}$$

$$\{[\text{Prob}(X \geq x_i^* | \pi_i)]^{-1} - 1\}^{-1},$$

where $Q = \text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)[\text{Prob}(X \geq x_i^* | \pi_i)]^{-1}$ is the value to be equated among subpopulations for selection fairness as specified by the Equal Probability Model. Again, suppose we have specified a minimum level of satisfactory criterion performance (y^*) and a constant conditional probability of success given selection (Q), then a predictor cut score x_i^* can be determined for each subpopulation π_i $(i = 1, \ldots, g)$. Given the values y^*, Q, and x_i^*, the conditional probability of failure given rejection ($\overline{Q}$) will be constant for each subpopulation π_i if the following condition is satisfied:

$$\{[\text{Prob}(X \geq x_i^*, Y \geq y^*|\pi_i) + \text{Prob}(X < x_i^*, Y < y^*|\pi_i)]$$

$$[\text{Prob}(X \geq x_i^*|\pi_i)]^{-1} - Q\}\{[\text{Prob}(X \geq x_i^*|\pi_i)]^{-1} - 1\}^{-1}$$

$$= \{[\text{Prob}(X \geq x_j^*, Y \geq y^*|\pi_j) + \text{Prob}(X < x_j^*, Y < y^*|\pi_j)]$$

$$[\text{Prob}(X \geq x_j^*|\pi_j)]^{-1} - Q\}\{[\text{Prob}(X \geq x_j^*|\pi_j)]^{-1} - 1\}^{-1}$$

for $i, j = 1, \ldots, g$.

This condition will be satisfied if and only if $\text{Prob}(X \geq x_i^*|\pi_i) = \text{Prob}(X \geq x_j^*|\pi_j)$ <u>and</u> $\text{Prob}(X < x_i^*, Y < y^*|\pi_i) = \text{Prob}(X < x_j^*, Y < y^*|\pi_j)$ for $i, j = 1, \ldots, g$, but not generally. In that case, the same selection strategy, the same set of predictor cut scores x_i^* $(i = 1, \ldots, g)$, is considered fair according to both the Equal Probability Model and the Converse Equal Probability Model, but, otherwise, the selection strategies will differ.

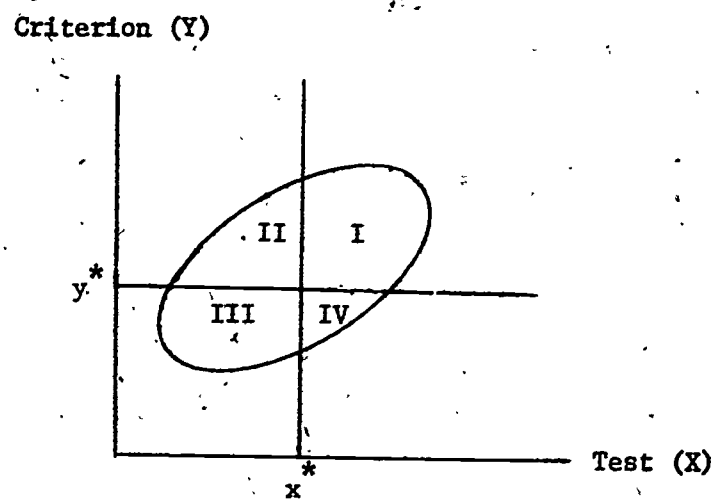
Figure 5 compares the Constant Ratio Model, the Conditional Probability Model, the Equal Probability Model and the three "converse" models. Representing proportions as areas, we see the six ratios for the six models. Unfortunately, there are indeed six models, each seemingly attractive and each paired with an indistinguishable converse. Which is the appropriate model? Or, indeed, is any one of them acceptable in any common situation?

The Equal Risk Model

Einhorn and Bass (1971) proposed a model for culture-fair selection which takes into account, for each subpopulation, the probability of success associated with an applicant's test score

Figure 5

A Comparison of Six Models for Culture-fair Selection



The cut score on the test (x^*) is determined so that the ratio (as specified by a particular model) is the same for all subpopulations.

Model	Ratio
Constant Ratio	(I + IV)/(I + II)
Conditional Probability	I/(I + II)
Equal Probability	I/(I + IV)
Converse Constant Ratio	(III + II)/(III + IV)
Converse Conditional Probability	III/(III + IV)
Converse Equal Probability	III/(III + II)

rather than just the applicant's predicted criterion score as suggested by the Regression Model. Their model is based on a definition of selection bias given by Guion (1966). Guion stated that

> unfair [test] discrimination exists when persons with equal probabilities of success on the job have unequal probabilities of being hired for the job. [p. 26.]

The objective of this model is not simply to accept those persons who are predicted, in the sense of best point estimate, to be above a specified minimum point on the criterion, but rather to accept those persons for whom this prediction can be made with a specified degree of confidence. The problem then becomes one of finding a cut score on the predictor variable so that the criterion score for persons with test scores greater than the cut score will be above the minimum acceptable criterion score with probability at least equal to some specified value. Furthermore, this model specifies that this probability (or, conversely, risk) must be the same in all subpopulations; hence the reference to it as the Equal Risk Model. Therefore, symbolically, at the minimum level of satisfactory criterion performance (y^*), the Equal Risk Model requires that the predictor cut scores x_i^* ($i = 1, \ldots, g$) be determined so that

$$Z = \text{Prob}(Y \geq y^* | X = x_1^*, \pi_1) = \ldots = \text{Prob}(Y \geq y^* | X = x_g^*, \pi_g) , \quad (8)$$

where Z is a fixed <u>constant</u> probability of success for all subpopulations π_i.

To illustrate, again suppose the applicants to an institution can be subdivided into two subpopulations, π_1 and π_2. Refer to Figure 6. (Adapted from Einhorn and Bass, 1971, pp. 265, 267.) Figure 6(a) shows the relationship between a predictor (test)

Figure 6

Illustration of Culture-fair Selection

as Defined by the Equal Risk Model

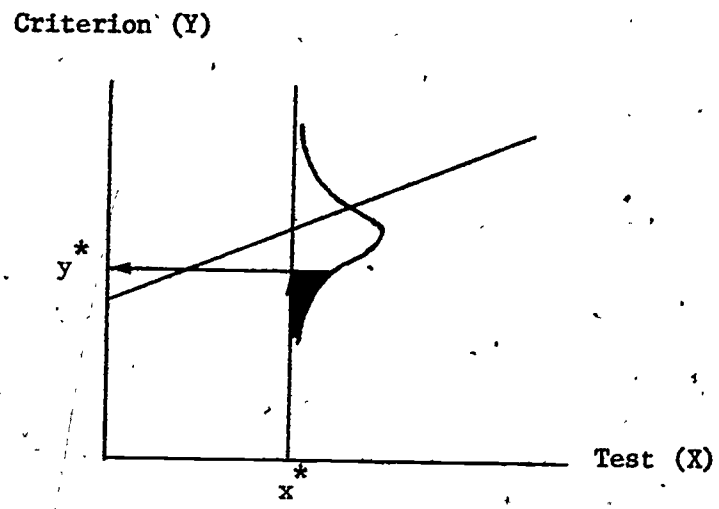


Figure 6(a). Conditional distribution of criterion on test showing risk level.

Figure 6 (cont'd.)

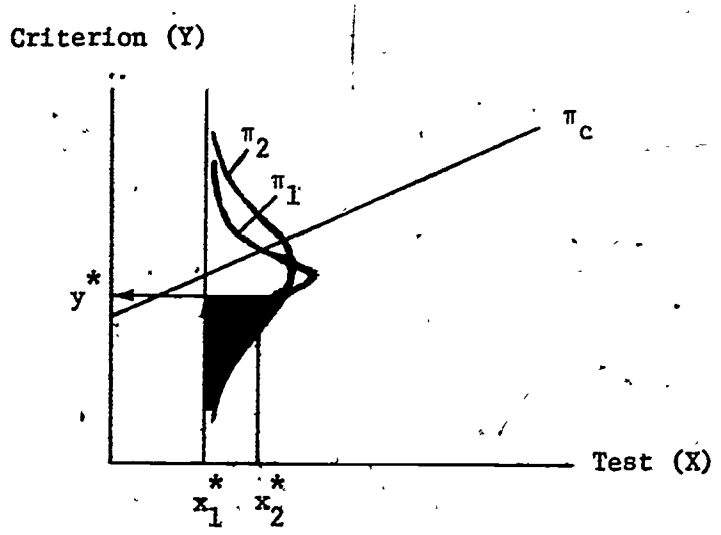


Figure 6(b). Subpopulations with common regression line but different standard errors of estimate.

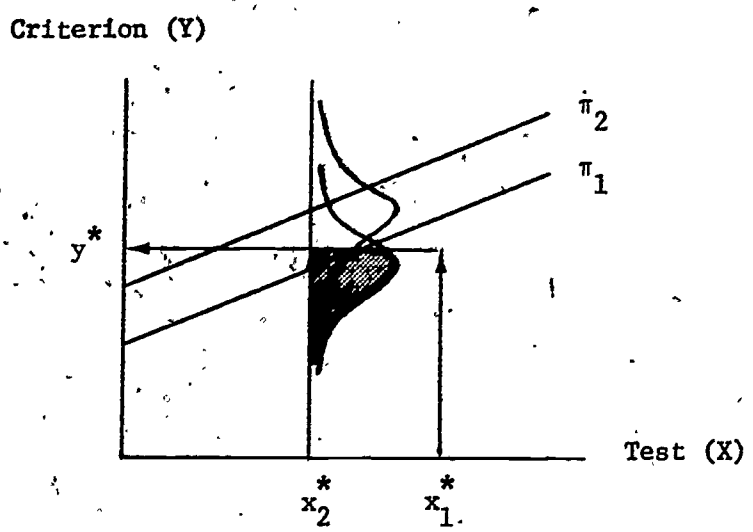


Figure 6(c). Subpopulations with the same standard error of estimate and the same slope but different intercepts.

variable and a criterion variable for one subpopulation. The conditional distribution of Y (criterion) given X (predictor) is assumed to be normal. The shaded portion of the distribution represents the risk level for a particular value x on the test. In Figure 6(b), the regression lines for the two subpopulations coincide; however, the standard error of estimate is smaller for subpopulation π_1 than for π_2. Provided, as in the figure, $y^* < y_.$ (the sample mean), then for any test score x, the level of risk is less for members of subpopulation π_1 than for members of subpopulation π_2. Thus, if all applicants with predicted criterion scores greater than or equal to y^* ($X \geq x_1^*$) were selected, the test would discriminate against subpopulation π_1 according to the Equal Risk Model. In this situation, to make the selection procedure fair (according to the Equal Risk Model), members of subpopulation π_1 with test scores greater than or equal to x_1^* would be accepted and members of subpopulation π_2 with test scores greater than or equal to x_2^* would be accepted. However, if the standard error of estimate had been the same in each subpopulation or if $y^* = y_.$, then the use of a single cut score would be considered fair to members of both subpopulations. In Figure 6(c), the two subpopulations have the same standard error of estimate and the same slope but different intercepts. For any test score x, the level of risk is less for a person from subpopulation π_2 than for a person from subpopulation π_1. If a single cut score is used, then the test discriminates against members of subpopulation π_2 according to the Equal Risk Model. The selection procedure would be considered fair (according to the Equal Risk Model) if members of subpopulation π_1

(π_2) with test scores greater than or equal to x_1^* (x_2^*) are accepted. Note that if each subpopulation has the same standard error of estimate and the same slope, then the selection strategies proposed by the Regression Model and the Equal Risk Model are the same. The two models are very closely related, differing only in that one is based on the mean-squared-error criterion and the other on a threshold-loss criterion, and both are straightforward applications of statistical decision theory.

The converse of the Equal Risk Model would require, given a minimum level of satisfactory criterion performance (y^*) that the predictor cut scores x_i^* ($i = 1, \ldots, g$) be determined so that

$$\bar{Z} = \text{Prob}(Y < y^* | X = x_1^*, \pi_1) = \ldots = \text{Prob}(Y < y^* | X = x_g^*, \pi_g) , \tag{9}$$

where $\bar{Z}$ is a fixed <u>constant</u> degree of risk for all subpopulations π_i. This relationship can be rewritten as

$$\bar{Z} = 1 - \text{Prob}(Y \geq y^* | X = x_i^*, \pi_i)$$

$$= 1 - Z ,$$

where $Z = \text{Prob}(Y \geq y^* | X = x_i^*, \pi_i)$ is the value to be equated among subpopulations for selection fairness as specified by the Equal Risk Model. Thus, the criterion of the Converse Equal Risk Model is a linear function of that of the Equal Risk Model. Hence, unlike the Constant Ratio Model, the Conditional Probability Model and the Equal Probability Model, the Equal Risk Model and its converse will always specify the same selection strategy. There is no internal contradiction; the model is coherent.

A Critique of the Constant Ratio, the Conditional Probability and the Equal Probability Models

One problem with the Conditional Probability Model and the Converse Conditional Probability Model is that each model treats only one aspect (selection-success) of the culture-fair selection issue. Recall that $K = \text{Prob}(X \geq x_i^* | Y \geq y^*, \pi_i)$ and $\overline{K} = \text{Prob}(X < x_i^* | Y < y^*, \pi_i)$ are the values to be equated among subpopulations for selection fairness as specified by the Conditional Probability Model and Converse Conditional Probability Model, respectively. In practice, we must consider equating both K and $\overline{K}$ among subpopulations. Since it can be shown that only under certain special conditions equating K among subpopulations leads to equating $\overline{K}$ among subpopulations, and vice versa (refer to the section entitled The Converse Conditional Probability Model), it might be suggested that in order to take both aspects of the culture-fair selection issue [the conditional probability of selection given success (K) and the conditional probability of rejection given failure ($\overline{K}$)] into consideration, we should at least contemplate equating some combination of K and $\overline{K}$ instead of trying to equate, independently, either K or $\overline{K}$ among subpopulations. However, it will be difficult to decide what function of K and $\overline{K}$ should be equated among subpopulations for fair test use.

Similar comments can be made regarding the Constant Ratio and Converse Constant Ratio models, and regarding the Equal Probability and Converse Equal Probability models. Each model deals with only one aspect of the culture-fair selection issue. In contrast, the definition of

culture-fair selection proposed by the Equal Risk Model deals with both sides of the issue, because if one equates Z [Equation (8)] among subpopulations, then one also equates the converse $\overline{Z} = 1 - Z$ [Equation (9)] among subpopulations.

To see why one should consider both aspects of the issue of fairness, note that if one tries to increase the conditional probability of selection given success (K), then one will decrease the conditional probability of rejection given failure ($\overline{K}$). Rewrite K and $\overline{K}$ as follows:

$$K = \frac{\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)}{\text{Prob}(Y \geq y^* | \pi_i)}$$

and

$$\overline{K} = \frac{\text{Prob}(X < x_i^*, Y < y^* | \pi_i)}{\text{Prob}(Y < y^* | \pi_i)}$$

Now, for a specified minimum level of criterion performance (y^*), $\text{Prob}(Y \geq y^* | \pi_i)$ and $\text{Prob}(Y < y^* | \pi_i)$ are fixed values. Thus, for K to increase, $\text{Prob}(X \geq x_i^*, Y \geq y^* | \pi_i)$ must increase implying that the predictor cut score (x_i^*) must decrease. (See Figure 3.) It is then clear that if x_i^* decreases, $\text{Prob}(X < x_i^*, Y < y^* | \pi_i)$ must decrease implying that $\overline{K}$ must decrease. Hence, although both large K and large $\overline{K}$ seem desirable for a given subpopulation π_i, any predictor cut score x_i^* which leads to an increment in K will result in a decrement of $\overline{K}$. This is similar to the situation in hypothesis testing where one tries to avoid two types of errors and, therefore, has to reach a compromise in selecting a critical region. Thus, if

one is inclined to build a model around Cole's conception of selection fairness, then one must try to equate a function of K and $\overline{K}$ among subpopulations rather than to equate K or $\overline{K}$ alone. To do this would require a value specification for the relative size of K and $\overline{K}$.

One can also show that in the case of the Constant Ratio Model and the Converse Constant Ratio Model, $\overline{R}$ [Equation (5)] will decrease as R [Equation (2)] increases. This indicates the same dilemma of trying to compromise between equating R or equating $\overline{R}$ among subpopulations. Thus, a definition of selection fairness can only be satisfactory if one considers both R and $\overline{R}$. Thus, among the Constant Ratio, the Conditional Probability, the Equal Probability and the Equal Risk models, only the Equal Risk Model is satisfactory in the sense that it takes both sides (selection-success and rejection-failure) of the culture-fair selection issue into account.

The Culture-Modified Criterion Model

In addition to the criterion variable Y and the predictor variable X, Darlington (1971) defines a third variable C, which denotes an applicant's group membership. The variable C may be either dichotomous or continuous (e.g., sex; race; socio-economic status). Darlington then gives (and discards) four definitions of

cultural fairness in terms of the correlations among the three variables X, Y, and C.

In order to state the four definitions in common correlational terminology, simplifying assumptions are introduced: the variables X and Y have a bivariate normal distribution in each subpopulation; the correlation between X and Y (r_{xy}) is positive; and the standard deviation on the test (s_x), the standard deviation on the criterion (s_y), and r_{xy} are constant for each subpopulation. Darlington's four definitions of cultural fairness are:

(1) $r_{cx} = r_{cy}/r_{xy}$,

(2) $r_{cx} = r_{cy}$,

(3) $r_{cx} = r_{cy}r_{xy}$, and

(4) $r_{cx} = 0$,

where the r's represent the correlations between the subscripted variables. In each case, a test is considered culturally fair if it satisfies the appropriate equation. (Darlington, 1971, p. 73.)

Definition (1) is equivalent to the Regression Model which requires a common regression line. Definition (2) is the same as Thorndike's Constant Ratio Model. Definition (3) is a special case of Cole's Conditional Probability Model. Definition (4) is the same as the requirement that subpopulations have equal means on the test. (Darlington, 1971, pp. 73-75; Linn, 1973, pp. 156-157.)

The four definitions yield contradictory results except in the case of perfect validity ($r_{xy} = 1$) or in the case of equal subpopulation means on the criterion ($r_{cy} = 0$). Darlington also claims that the four definitions are

> all based on the false view that optimum treatment of cultural factors in test construction or test selection can be reduced to completely mechanical procedures. If a conflict arises between the two goals of maximizing a test's validity and minimizing the test's discrimination against certain cultural groups, then a subjective, policy-level decision must be made concerning the relative importance of the two goals. [p. 71.]

Darlington then suggests that instead of predicting the criterion variable Y that a variable (Y - kC) be defined where k is determined by a subjective value judgment on the part of the decision maker (test user). Darlington urges that

> the term "cultural fairness" be replaced in public discussions by the concept of "cultural optimality." The question of whether a test is culturally optimum can be divided in two: a subjective, policy-level question concerning the optimum balance between criterion performance and cultural factors (operationalized ... as the optimum value of k), and a purely empirical question concerning the test's correlation with the culture-modified variable (Y - kC) and whether that correlation can be raised. [pp. 79-80.]

According to this formulation, each institution must first choose a value of k, indicating whether there is special value in the selection of members from some subpopulation. That is, the decision maker must answer the question, "How many units on Y are considered equivalent in value to one unit on C?". Then, the psychometrician's job is to contruct a test to predict the variable (Y - kC). Note that when k is set equal to zero, that is, when there is no reason to favor one cultural group, this procedure reduces to that of the Regression Model. Also

note that where the other models for culture-fair selection would set different predictor cut scores for each subpopulation, Darlington would add a specified number of points to the scores of one subpopulation and then use the same predictor cut score.

Darlington's formulation of selection fairness recognizes, explicitly, that the variable which is traditionally considered to be the criterion (e.g., college grade point average) is not the only criterion. Group membership or culture is also part of the criterion. Darlington, then, argues that the traditionally accepted criterion must be modified for culture; hence the reference to Darlington's formulation of selection fairness as the Culture-Modified Criterion Model.

An Appraisal of the Test Bias Models

The Regression, the Constant Ratio, the Conditional Probability, the Equal Probability, the Equal Risk and the Culture-Modified Criterion models are each explications of general concepts of what constitutes the fair use of tests in a selection situation. There seems to be nothing in the literature that clearly indicates when, if ever, one of the models is preferable to the other five models. Thus, the practitioner has no clear guidance in the choice of a culture-fair selection model. Further, we have suggested that the Constant Ratio, the Conditional Probability, the Equal Probability models and their converses are internally contradictory.

There has been considerable interest in the Constant Ratio Model and the Conditional Probability Model based on the fact that these models yield a popular result, in that they apparently give

lower cut scores for disadvantaged minority populations. The appeal of these models, then, is that they produce a desirable result.

However, we might well contend that it is generally not appropriate to evaluate the correctness of a model solely on the basis of the pleasantness or unpleasantness of its implications but, rather, that one must look carefully at the logical structure of the model. One must be sure that the model is getting the right results for the right reasons. If the models are giving the right results for the wrong reasons, it may well be possible that, in some other circumstances, wrong answers will be forthcoming. We shall see that this is indeed the case. We shall show that these models sometimes can produce most undesirable results and could, in fact, be used to justify discrimination against some minority groups.

To see that this may happen, consider a situation in which the regression lines in the minority and the majority populations are identical, but in which the mean values of X and Y are higher in the minority (disadvantaged) population and lower in the majority (advantaged) population. (Refer to Figure 4.). This situation is not typical but, in fact, can be found if one compares, for instance, a Japanese-American minority population with the Anglo majority population. In this situation, both the Constant Ratio and the Conditional Probability models will give lower predictor cut scores and hence easier entry to the majority population. The Regression Model and the Equal Risk Model will give identical predictor cut scores. If, as well may be the case, the Japanese-American subpopulation has been discriminated against in some situations and is thus a disadvantaged group, then our desire might be to provide easier access

for that subpopulation, but, in fact, the two models being considered make access more difficult.

From this example, it can be seen that the two models being discussed make a correction that is usually in the desirable direction, but that they make that correction for the wrong reason. They make the correction simply because of differences in the mean values of X and Y in the two populations and they only coincidentally take into account the public desire or social necessity to rectify unfair treatment to a minority population. The degree of advantage to the minority population is largely a function of the difference in the mean test scores and in no way directly reflects the degree of prior discrimination and disadvantage that group has suffered. On the other hand, if, following the general ideas to be laid down here, one allows that differential treatment should be given to a degree agreed upon by the political process to some heretofore disadvantaged group, then a lower predictor cut score will be obtained for that group. In this case, the lower score is obtained for the right reason, because of their disadvantaged status (different utility structure), and not simply because of a difference in mean values. We judge that for these reasons the use of the Constant Ratio, Equal Probability and Conditional Probability models and their converses is contraindicated. In stating this, it is not suggested that any ill effects will necessarily result from their use. Only by more detailed study would it be possible to document more completely all of the situations in which these models break down. However one could easily name other minority groups, discrimination against which would be sanctioned by any official, or other, recommendation, endorsement or adoption of any of these models. Finally, we would remark that there is no reason to believe that either the Conditional Probability or Constant Ratio models will provide, under any circumstances, the degree of compensation that I think certain disadvantaged persons should have.

Maximizing Expected Utility

There exists a body of quantitative reasoning, whose origins are ancient and remote, that has received codification in this century in the work of Von Neumann and Morgenstern (1947), Wald (1950), and others. In the theory of the rational-economic man, developed in these writings, when all probabilities of outcomes are assumed known (an assumption made explicitly here and implicity in previous statements of the models under consideration) there is a simple paradigm required for rational decision. In that paradigm the desirability or utility of each possible outcome is stated quantitatively. Then, given all available information concerning the person in question, the probability of each possible outcome is stated for each decision under consideration. Next, for each possible decision, the utility of each outcome is multiplied by the probability of each outcome and the products are summed to provide an expected utility. Finally, that decision is then made for which the expected utility is highest. Most statisticians interested in decision problems accept the correctness of the Von Neumann and Morgenstern-Wald model and the incorrectness of any statistical decision procedure that does not conform to that model. It seems clear that the Constant Ratio Model, the Conditional Probability Model and the Equal Probability Model do not conform to that model, though the ideas that are at their bases may well be reformulated in a coherent manner.

If the utility of an outcome depends only on the individual outcome (success-failure) and the subpopulation involved, then the

axioms of rational decision-making require the maximization of expected utility (but see Suppes, 1974, for a slight weakening of this statement that is not relevant here). That maximization process involves only the conditional probability of success given test score and the utility of that success. It does not involve any of the marginal or conditional probabilities used by Thorndike, Cole or Linn. From this point of view, the fundamental fallacy in each of these models is that they are based on the wrong conditional probability.

Specifically, the conditioning process must be on the specific value x observed on the person and not on the marginal distribution of X (Thorndike), the conditional distribution of X given y (Cole) or the event $X \geq x$ (Linn). The three models mentioned above do not use the correct probability. The Regression Model and the Equal Risk Model do, and it is for this reason that no logical contradictions have arisen with these models. This is not to say that these latter models are entirely satisfactory; indeed, You could judge them to be generally unsatisfactory but only because You judge the utilities they adopt to be inappropriate. While these models are both special cases in the general decision-theoretic formulation they are, it would seem, much too special. The utilities they adopt, implicitly, are not particularly compelling and unfortunately they do not make these assumptions very clear. More general models are required.

Some individuals (e.g., Humphreys, 1973) have indicated a basic dislike of differential treatment of groups while possibly accepting its short-term desirability. That position has merit,

though in the current climate of opinion, it may represent a minority view. One coherent expected utility model (the Equal Risk Model) provides for equal treatment of all persons. However, we suggest that if this criterion is appropriate it would be better to arrive at it on the basis of careful analysis and debate in the area of public policy rather than because of some notion regarding the universal applicability of any one model.

Thorndike has argued forcefully that the marginal distributions of both X and Y are important in culture-fair testing whereas the decision-theoretic formulation we discuss in the next section concentrates only on the conditional distribution of Y given x. (Thorndike's view, which is that some consideration must be given in the setting of cutting scores to their effect on the percentage of successful persons in each subpopulation, can possibly be accommodated within a decision-theoretic framework.) It would seem to us that it <u>might</u> be appropriate in assessing utilities in the two subpopulations to take into consideration the implications with respect to these marginal distributions. We would expect, however, this consideration of marginal distributions to also take into account the effect on the percentage of failures in the two subpopulations. Such investigation could result in our utilities being related to the location parameters of the marginal distributions, but this would not affect the <u>probability</u> aspect of the decision-theoretic formulation which would still depend only on the distribution of Y given x. A similar remark might be made with respect to Cole's conception of culture-fair selection which might be reformulated in terms of utilities rather than probabilities. However, in our judgment, until

these ideas are given a strict decision-theoretic formulation in terms of utilities that You and I can understand, we must caution against any use, adoption or recommendation of these ideas.

Darlington's Culture-Modified Criterion Model is the only model surveyed that addresses itself to the utility question. It also has the desirable feature of focusing on the correct conditional probability. Unfortunately, this formulation is still not entirely consistent with the decision-theoretic approach (i.e., it does not incorporate a formal utility function), and hence may not be acceptable. We would urge Darlington to restate his model so that it can be evaluated more easily.

The Threshold Utility Model

In a recent paper, Gross and Su (1975) investigated one decision-theoretic approach to the culture-fair selection problem. This formulation maximizes expected utility using a threshold utility model. The discussion in this section is taken from a somewhat more complete analyses due to Petersen (1974).

Assume that the applicants to an institution can be separated into two subpopulations referred to as the minority population (π_1) and the majority population (π_2). Further, assume that a predictor variable X (a test score) and a criterion variable Y (a measure of performance) have possibly different (unspecified) positively correlated bivariate distributions in each subpopulation.

Each applicant to the institution will, if accepted, be either a successful performer or an unsuccessful performer. The state of the applicant being a successful performer is represented by those values of Y greater than or equal to the point y^*, and the

state of the applicant being an unsuccessful performer is represented by those values of Y less than the point y^* where y^* is the minimum level of satisfactory performance on the criterion variable Y. Two possible actions are open to the institution. The institution can either accept the applicant or reject the applicant. The accept decision is represented by those values of X which are greater than or equal to the point x_i^*, and the reject decision by those values of X which are less than the point x_i^* where x_i^* is the cut score on the predictor variable X_i and may differ for each subpopulation π_i, i = 1, 2.

For each subpopulation π_i, the action decided upon by the institution can have one of four outcomes:

O_1: $X \geq x_i^*$ $Y \geq y^*$ An applicant is accepted and is successful.

O_2: $X < x_i^*$ $Y \geq y^*$ An applicant is rejected but would have been successful.

O_3: $X < x_i^*$ $Y < y^*$ An applicant is rejected and would have been unsuccessful.

O_4: $X \geq x_i^*$ $Y < y^*$ An applicant is accepted but is unsuccessful.

Outcomes O_1 and O_3 are desirable since they represent correct decisions, whereas outcomes O_2 and O_4 are undesirable since they represent incorrect decisions. Up to this point the formulation is the same as for the constant ratio, conditional probability, equal probability and equal risk models.

With these outcomes in mind, the institution designates a threshold utility function $u(O_j|\pi_i)$, j = 1, 2, 3, 4, for each subpopulation π_i, i = 1, 2, as defined in Figure 7. For the

Figure 7

Threshold Utility Function

$u(O_j|\pi_i)$

	$X < x_i^*$ (reject)	$X \geq x_i^*$ (accept)
$Y \geq y^*$ (successful)	O_2 b_i	O_1 a_i
$Y < y^*$ (unsuccessful)	O_3 c_i	O_4 d_i

subpopulation π_i, the utilities $a_i = u(O_1|\pi_i)$ and $c_i = u(O_3|\pi_i)$ are associated with correct decisions and should be larger than the utilities $b_i = u(O_2|\pi_i)$ and $d_i = u(O_4|\pi_i)$. The zero point is arbitrary, though it may be convenient to take b_i, $d_i \leq 0$ and refer to these as disutilities. The degree of preference given any group will depend upon the utilities for that group.

Regardless of the ordering of the utilities, the expected utility of selection for an applicant from subpopulation π_i is given by

$$\mathcal{E}[u(O|\pi_i)] = \sum_{j=1}^{4} u(O_j|\pi_i)\text{Prob}(O_j|\pi_i)$$

$$= a_i[\text{Prob}(O_1|\pi_i)] + b_i[\text{Prob}(O_2|\pi_i)]$$

$$+ c_i[\text{Prob}(O_3|\pi_i)] + d_i[\text{Prob}(O_4|\pi_i)] \quad (1)$$

Now, the process of selection is viewed as a series of separate decisions, each of which involves one applicant. Thus, the expected utility of the selection process is found by summing the expected utility of selection for an applicant [Equation (1)] over all applicants, that is,

$$\mathcal{E}[u(O)] = \sum_{i=1}^{2} p_i\{\mathcal{E}[u(O|\pi_i)]\}$$

$$= \sum_{i=1}^{2} p_i \sum_{j=1}^{4} u(O_j|\pi_i)\text{Prob}(O_j|\pi_i) ,$$

where p_i is the proportion of the combined applicant population (π_1 and π_2) who are members of subpopulation π_i. The problem is to find predictor cut scores x_1^* and x_2^* such that the expected utility of the

selection process [Equation (10)] is a maximum. The method for doing this is given by Gross and Su and by Petersen.

In effect what this model does is to select those applicants with the highest expected utility selection. The utilities will typically differ for different subpopulations. (If they do not, we have the Equal Risk Model.) The prediction equations may also differ in the subpopulations.

The advantage of this model is that it requires an explicit public statement of utilities for each subpopulation. If I do not like the utilities You provide, an informative public debate will no doubt ensue, with neither of us claiming any axiomatic justification for our utilities. The important thing is that the discussion be public with all interested parties participating in the debate.

In our judgment, the concepts of culture-fairness and group parity are neither useful nor tenable, and the models spawned from them should not enjoy institutional endorsement. The problem, we think, should be reconceptualized as a problem in maximizing expected utility. The Threshold Utility model is one possibly useful model. It has its limitations. The concerns that motivated the Thorndike, Cole and Linn models are important ones; however they can be explicated in conceptually simple extensions of the Threshold Utility model. This might involve dropping the additivity assumption of this model. Our main purpose in this paper has been to show that the ideas of culture-fairness and group parity have spawned incoherent models that can sanction the very discrimination they seek to rectify. Any new explications of these ideas will need to be scrutinized carefully from statistical, psychological, social, ethical and legal points of view.

References

ANASTASI, A. Psychological Testing. (3rd Ed.) New York: Macmillan, 1968.

BARTLETT, C. J., & O'LEARY, B. S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification. Personnel Psychology, 1969, 22, 1-17.

BOWERS, J. The comparison of GPA regression equations for regularly admitted and disadvantaged freshmen at the University of Illinois. Journal of Educational Measurement, 1970, 7, 219-225.

CLEARLY, T. A. Test Bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

COLE, N. S. Bias in selection. Journal of Educational Measurement, 1973, 10, 237-255.

CRONBACH, L. J., & GLESER, G. C. Psychological Tests and Personnel Decisions. (2nd Ed.) Urbana: University of Illinois Press, 1965.

DARLINGTON, R. B. Another look at "cultural fairness". Journal of Educational Measurement, 1971, 8, 71-82.

EINHORN, N. J., & BASS, A. R. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.

GROSS, A. L., & SU, WEN-HUEY. A decision theory approach to test bias. To appear in Journal of Applied Psychology.

GUION, R. Employment tests and discriminatory hiring. Industrial Relations, 1966, 5, 20-37.

HUMPHREYS, LLOYD G. Implications of group differences for test interpretation. Assessment in a Pluralistic Society: Proceedings of the 1972 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1972.

LINN, R. L. Fair test use in selection. Review of Educational Research, 1973, 43, 139-161.

LINN, R. L., & WERTS, C. E. Considerations for studies of test bias. Journal of Educational Measurement, 1971, 8, 1-4.

PETERSEN, NANCY S. An expected utility model for "optimal" selection. Doctoral Dissertation. University of Iowa, 1974.

SCHMIDT, F. L., & HUNTER, J. E. Racial and ethnic bias in psychological tests. American Psychologist, 1974, 1-8.

SUPPES, P. The measurement of belief. Journal of the Royal Statistical Society, Series B (Methodological), Volume 36, 2, 1974.

TEMP, G. Validity of the SAT for blacks and whites in thirteen integrated institutions. Journal of Educational Measurement, 1971, 8, 245-251.

THORNDIKE, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

VON NEUMANN, J., & MORGENSTERN, O. Theory of Games and Economic Behavior. (2nd ed.) Princeton: Princeton University Press, 1947.

WALD, A. Statistical Decision Functions. New York: Wiley, 1950.