

The Cleary Model: "Test Bias" as Defined By the EEOC Uniform Guidelines on Employment Selection Procedures

The following section of the [EEOC Uniform Guidelines on Employee Selection Procedures](#) provides a definition of "test fairness:"

29 CFR Ch. XIV (7-1-99 Edition) (a) **Unfairness defined.** When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.



One of the most common ways to determine whether a test is "fair" involves what has come to be known as the "Cleary model" or "regression model." T. Anne Cleary (1968) first described this method over 30 years ago. Consider Figure 1 below in which personnel selection test scores X are plotted against subsequent measures of job performance Y.

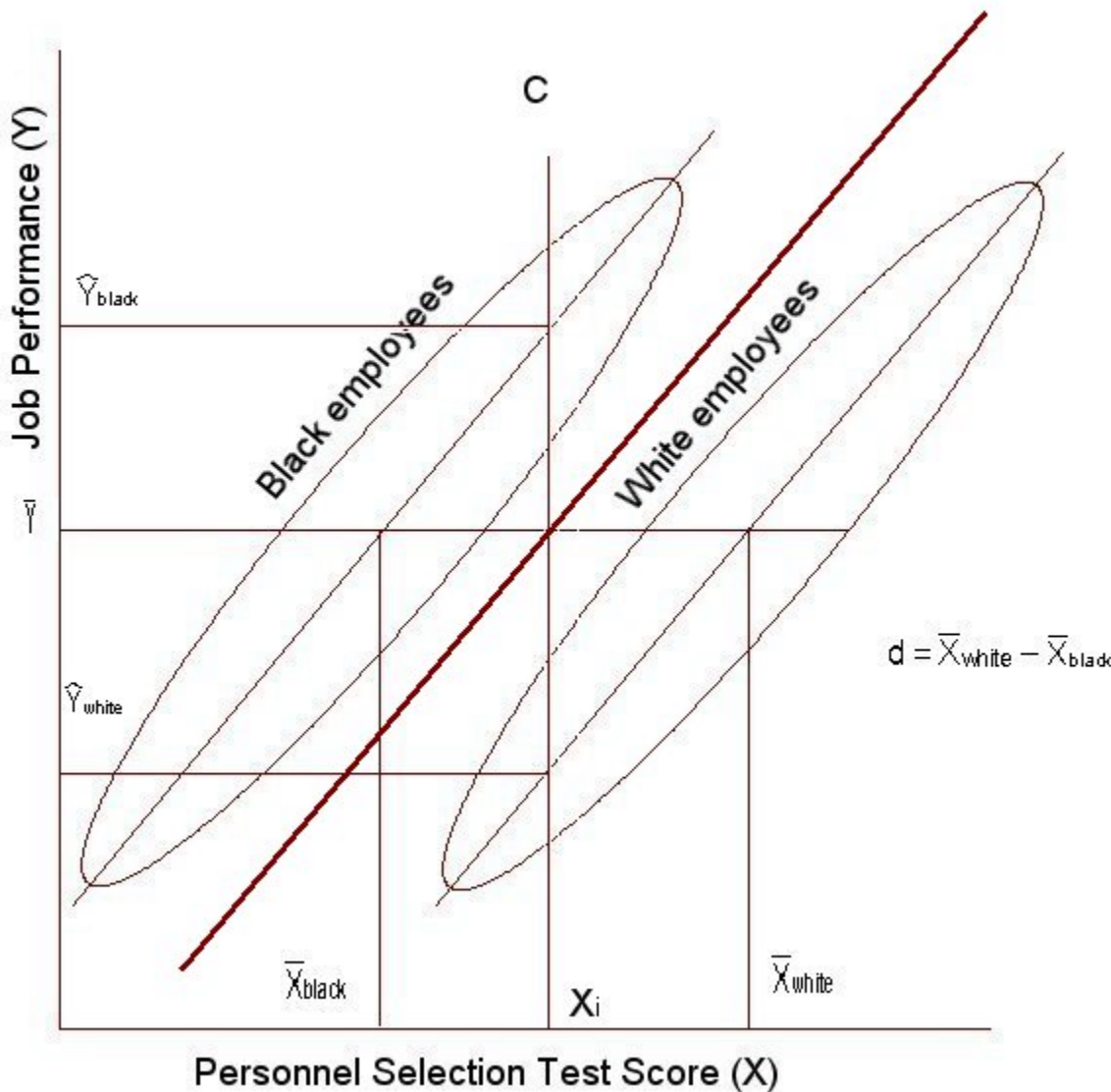


Figure 1: An Unfair Test (or two hot dogs on separate sticks).

Note the following points of interest:

1. Both black and white employees perform equally well on the job.
2. Black employees perform meaningfully lower on the personnel selection test relative to white employees (by the amount d).
3. For any "cut score" C (i.e., a vertical line drawn from the X axis upwards signifying the minimum X score needed to receive a job offer), more white applicants will receive job offers than blacks.
4. A black applicant who earns score X_i will be expected to generate job performance equal to \hat{Y}_{black} , while a white applicant who earns the same X_i score will be expected

Note, in the late 1980's my daughter wanted to see what I did at work. Consequently, I let her attend one of the masters level statistics classes I was teaching for the Rutgers School of Management and Labor Relations Human Resources degree program. That night I covered the Cleary model and, upon asking her how she liked the course, she complimented me on my "hot dogs and sticks." Hence, a biased test will be one in which the two hot dogs are not on the same stick! If this metaphor helps you understand the issue,

A Statistical Interlude

Skip this part if you want a conceptual understanding of test fairness and don't plan on actually crunching numbers to determine its presence or absence.

Figure 1 describes a classic example of an unfair or "biased" test. The preferred method for statistically detecting this circumstance is hierarchical moderated regression analysis. This procedure involves an F test of the following null hypothesis:

$H_0 : \Delta R^2 = 0$, where

$\Delta R^2 = R_{\text{mult}}^2 - R_{\text{add}}^2$, and

R_{add}^2 is for $\hat{Y} = b_0 + b_1X_1$,

R_{mult}^2 is for $\hat{Y} = b_0 + b_1X_1 + b_2X_2$,

Y = job performance,

X_1 = selection test score, and

X_2 = race (1 = white, 2 = black)

With adequate sample size (which can pose a problem when N_{blacks} is very small), circumstances like those found in Figure 1 will cause the F statistic;

$$F_{1,N-3} = \frac{\Delta R^2}{(1 - R_{\text{mult}}^2)/(N - 3)}$$

to reject the null hypothesis

$$H_0 : \Delta R^2 = R_{\text{mult}}^2 - R_{\text{add}}^2 = 0$$

Statistically, this is equivalent to testing the null hypothesis that there is no significant difference between the slopes and intercepts of the lines running through the white and black ellipses.

Clearly any application of a personnel selection system exhibiting X-Y relationships found in Figure 1 will cause adverse impact -- no matter where a cut score is set, more whites will be selected than blacks **and** the proportion of blacks hired relative to their availability in the applicant pool will be meaningfully smaller than the proportion of whites hired relative to their availability (i.e., the 4/5ths rule will likely be violated).

Figure 2 below describes an instance where a test does not exhibit bias (i.e., it **is** a fair test) yet still causes adverse impact.

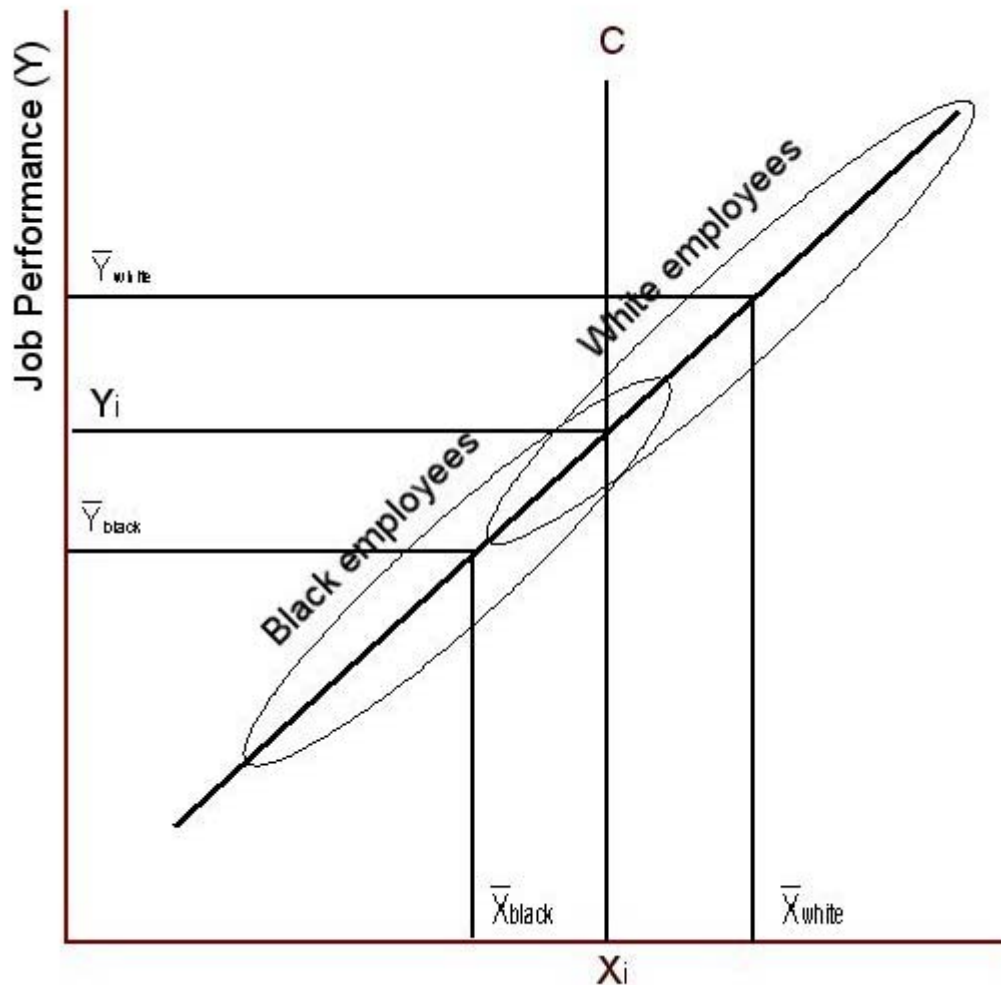


Figure 2: A Fair Test (or two hot dogs on a single stick).

Note the following points of interest:

1. Black and white applicants did not have the same average on the personnel selection test **or** subsequent job performance.
2. A black and white applicant with the same personnel selection test score X_i will be expected to generate the same level of job performance Y_i .
3. For any "cut score" **C** (i.e., a vertical line drawn from the X axis upwards signifying the minimum X score needed to receive a job offer), more white applicants will receive job offers than blacks. Hence, even though this personnel selection test is fair, it will still have adverse impact against blacks (no matter where a

Finally, Figure 3 shows a test that meets the EEO Guidelines fairness criterion yet exhibits differences in the **strength** of the X-Y relationship for black and white applicants. This was commonly referred to as "differential validity" during the 1970's, when conceptual definitions of test fairness were being thrashed out in the literature. See original research by Bartlett, Bobko, and Moser (1978) and Bobko and Bartlett (1978) for a sampling of studies addressing these issues. See Arvey and Faley (1992) for a complete discussion of these issues.

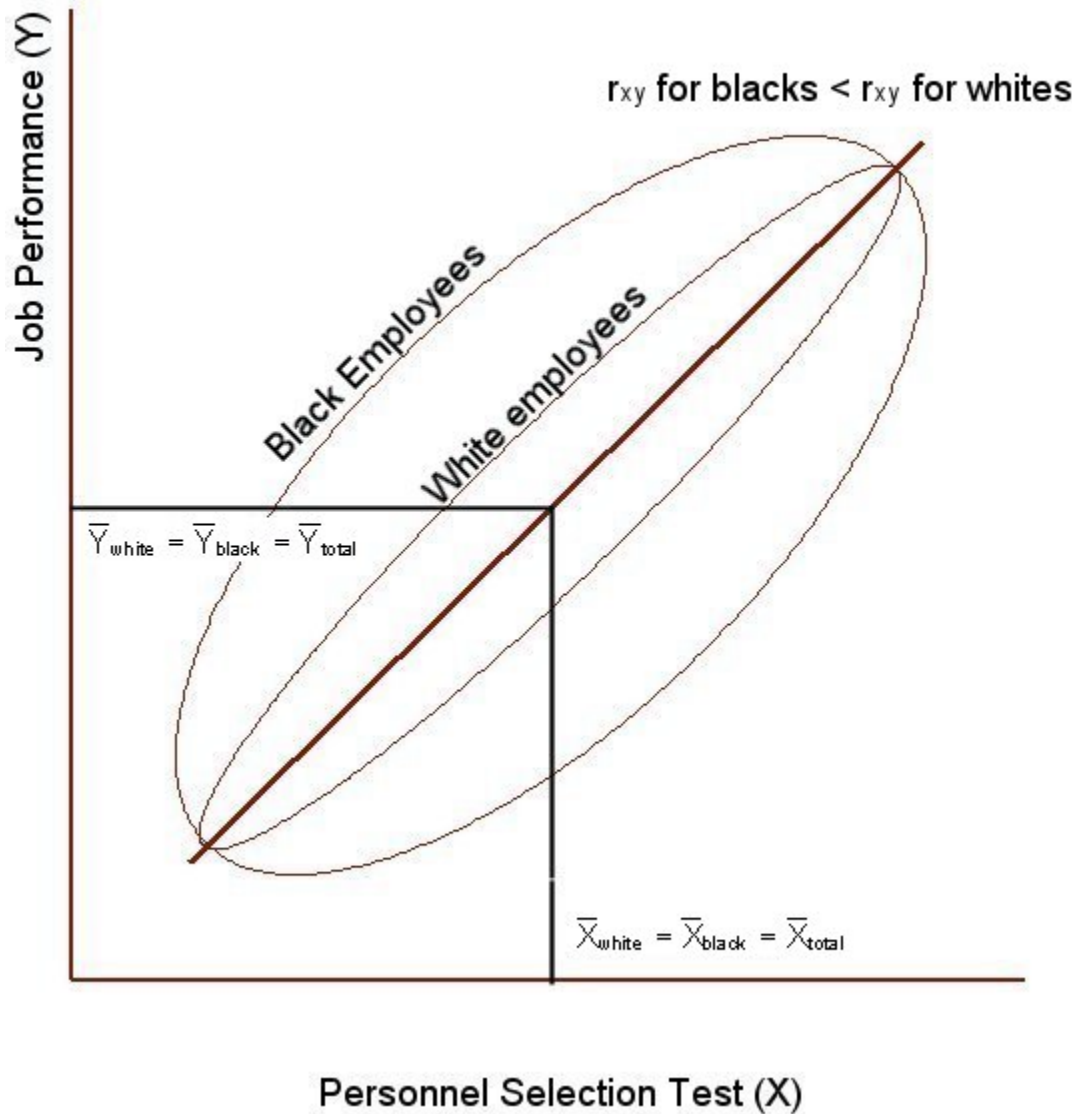


Figure 2: Differential Validity in a Fair Test (cheese dog.

Note the following points of interest:

1. Black and white mean job performance **AND** black and white mean selection test performance are equal.
2. No adverse impact will occur - no matter where a cut score is drawn, the proportion of whites hired relative to the number of whites applying is expected to be equal to the proportion of blacks hired relative to the number of blacks applying.
3. The predicted job performance for a black applicant and white applicant who earned the same selection test score X will be the same, though the accuracy

Continuing from the notes below Figures 1 and 2. this is an unbiased test because the two hot dogs are on the

same stick. Figure 3 is perhaps best conceived in terms of an Oscar Meyer product called a "cheese dog," i.e., a hot dog product (the outer ellipse) which has been injected with an inner "ellipse" of cheese.

of our prediction is greater for the white applicant.

It is important to note that inferences of test fairness are not related to whether the test exhibits equal (or comparable) criterion validity of black and white applicants (i.e., $r_{xy} = r_{xy}$ for blacks and whites). Perhaps the most perplexing aspect of "differential validity" is the frequently reported finding of:

$H_0 : r_{xy\text{whites}} = 0$, rejected

$H_0 : r_{xy\text{blacks}} = 0$, fail to reject,

$H_0 : r_{xy\text{whites}} - r_{xy\text{blacks}} = 0$, fail to reject.

References

Arevey, R.D. & Faley, R.H. (1992). *Fairness in selecting employees*. Reading, MA: Addison-Wesley.

Bartlett, C. J., Bobko, P., & Mosier, S. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology*, 31, 233-245.

Bobko, P. & Bartlett, C.J. (1978). Subgroup validities: Differential definitions and differential prediction. *Journal of Applied Psychology*, 63, 12-27.

Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.

© Craig J. Russell, 2000

[Back to White Pages](#)