HUMAN PERFORMANCE, *11*(4), 351–365

A New Procedure for Computing Equivalence Bands in Personnel Selection

Herman Aguinis
College of Business and Administration
University of Colorado at Denver

Jose M. Cortina
Department of Psychology
George Mason University

Edie Goldberg
Towers Perrin
San Francisco, California

We advance and illustrate a new procedure for the formation of equivalence bands in preemployment tests for situations in which criterion data are available. The procedure consists of 3 steps: (a) the computation of the width of a band of statistically indistinguishable scores on a performance measure *Y*, (b) the determination of the upper and lower limits on the band for *Y*, and (c) the computation of the 2 scores on a preemployment test *X* that produce predicted scores for *Y* equal to the upper and lower limits of the band on *Y* established in Step 2. Thus, this new approach generates 2 scores on *X* that are associated with a range of statistically indistinguishable scores on the predicted value of the criterion.

Personnel specialists are currently faced with a paradoxical situation: The use of cognitive abilities and other valid predictors of job performance leads to adverse impact (Schmidt, 1993). Thus, choosing predictors that maximize economic utility

Requests for reprints should be sent to Herman Aguinis, College of Business and Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217–3364. E-mail: haguinis@castle.cudenver.edu. URL: http://www.cudenver.edu/~haguinis

(as it is typically conceptualized in human resources management; Schmidt, 1991) often leads to the exclusion of members of protected groups (Sackett & Wilk, 1994).

Cascio, Outtz, Zedeck, and Goldstein (1991) proposed the use of preemployment test score banding as a way to incorporate both utility and adverse impact considerations in the personnel selection process. Banding is an alternative to the strict top-down selection strategy that typically leads to adverse impact, and is based on the premise that preemployment measures are never perfectly reliable. Thus, an observed difference in the scores of two job applicants may be the result of measurement error instead of actual differences in the construct that is measured. Consequently, if it cannot be determined with a reasonable amount of certainty that two applicants differ on the construct underlying a predictor score, there may be little reason to believe that they will differ with respect to subsequent job performance.

The Cascio et al. (1991) proposition regarding the use of equivalence bands in personnel selection has prompted a heated debate (Schmidt, 1991; Schmidt & Hunter, 1995) that has extended to the legal arena (Barrett, Doverspike, & Arthur, 1995; Gutman & Christiansen, 1997). However, these challenges have been contested in an equally vigorous manner (Cascio, Goldstein, Outtz, & Zedeck, 1995; Murphy & Myors, 1995; Zedeck, Outtz, Cascio, & Goldstein, 1991). Thus, the question of whether banding should be routinely implemented in personnel selection decision making is a highly politicized and emotional issue. However, as stated by others (e.g., Cascio et al., 1991), we believe that the use of banding may serve the dual purpose of fulfilling economic needs as well as achieving social goals. This article advances a new procedure for computing equivalence bands that will prove useful to personnel specialists predisposed to implement banding procedures in staffing decision making.

THE CASCIO ET AL. (1991) PROCEDURE FOR COMPUTING BAND WIDTH

To establish whether two observed scores are actually not reliably different (i.e., indistinguishable) regarding the underlying construct being measured by a preemployment test, it is necessary to compute bands of scores that take into account measurement error. Then, if two observed test scores fall within the same band, they are considered indistinguishable regarding the underlying construct of interest.

Based on the reliability estimate of the test, Cascio et al. (1991) proposed the following equation to compute band widths:

$$C \cdot s_x \cdot (1 - r_{xx})^{1/2} \cdot \sqrt{2}, \tag{1}$$

where C is the standard score indicating the desired level of confidence (e.g., 1.96 indicates a 95% confidence interval, 1.00 indicates a 68% confidence interval), s_x is the standard deviation of the test, and r_{xx} is an estimate of the internal consistency of the test measured on an interval-level scale. Substantively, $s_x \cdot (1 - r_{xx})^{1/2}$ is the standard error of measurement of the test and $s_x \cdot (1 - r_{xx})^{1/2} \cdot \sqrt{2}$ is the standard error of the difference (SED) between two scores on the test. If the test scores of two applicants fall within a band whose width is defined by Equation 1, it is concluded that these two applicants are indistinguishable with respect to the underlying predictor construct.

Because the scores within a band are considered to be indistinguishable, job applicants who score within the same band are considered equally qualified for the job in question. Choices can then be made among these "equivalent" applicants based on criteria other than test scores such as diversity considerations (Cascio et al., 1995). Thus, banding allows decision makers the flexibility to consider the diversity needs of the company and society as well as the standing of applicants with respect to the construct measured by the preemployment test.

THE NEED TO INCORPORATE CRITERION INFORMATION AND PURPOSE OF THIS STUDY

Banding is advocated as a procedure that considers "all scores falling within the band as relatively equally qualified, within the limits of measurement error" (Cascio et al., 1991, p. 242), but the equation used to calculate band width does not include specific criterion information. The ultimate goal of banding, and any other decision-making procedure in staffing situations, is to ascertain whether two applicants will perform at similar levels on the job. Hiring decisions are then made based on these predictions. However, the model advanced by Cascio et al. (cf. Equation 1) uses only information about the predictor to compute bands. Consequently, the only conclusion that follows logically from a band formed with the Cascio et al. procedure is that two applicants whose predictor scores fall within the same band do not differ regarding the construct underlying the preemployment test in use (e.g., cognitive abilities, conscientiousness). It is by inference that the Cascio et al. approach leads to the conclusion that these two applicants are unlikely to differ regarding job performance.

The Cascio et al. (1991) model does not explicitly consider the precise predictor–criterion relation, and operates under the assumption that there is an acceptable level of useful empirical or content validity. Accordingly, based on this "acceptable validity" premise, equivalence regarding predictor scores is equated with equivalence regarding criterion scores. However, few preemployment tests explain more than one fourth of the variance in a given criterion. Thus, the assumption that two applicants who are indistinguishable (i.e., falling within the same band) or distinguishable (i.e., not falling within the same band) regarding the predictor construct are also indistinguishable or distinguishable regarding the criterion construct may not be tenable.

Personnel specialists need to estimate whether applicants are likely to differ to a practically meaningful degree on the performance construct based on their observed predictor scores. The procedure that is presently available for computing band width does not allow such specific estimation. Two applicants falling within the same band on a given predictor will show similar performance levels as long as the validity of the test is very high. However, as the validity coefficient approaches the typical .20 to .50 range, the rank order of applicants with respect to predictor scores can be very different from their rank order with respect to criterion scores, and applicants who fall within the same band on the predictor can easily have dissimilar levels of performance. The Cascio et al. (1991) banding model includes neither specific validity information nor criterion information in computing band width. As a consequence, there is uncertainty regarding whether two applicants falling within the same band on the predictor will show equivalent levels of performance. The degree of uncertainty increases as the validity coefficient decreases.

This article advances a new approach for the computation of bands that allows personnel specialists to decide whether two applicants with different observed scores on a preemployment test are predicted to show reliable differences in the primary variable of interest; namely job performance. The proposed banding method can be easily implemented in situations in which criterion data are available. The sections that follow (a) describe this new approach in detail, (b) illustrate its implementation using data from a published article, (c) show the impact of variations in reliability and validity values on the obtained band width, and (d) compare band widths obtained using the new approach with band widths obtained using the Cascio et al. (1991) model.

A THREE-STEP PROCEDURE

The new approach for computing equivalence bands involves three steps: (a) the computation of the width of a band of statistically indistinguishable scores on a performance measure Y, (b) the determination of the upper and lower limits on the band for Y, and (c) the computation of the two scores on a preemployment test X that predict scores for Y equal to the upper and lower limits of the band on Y established in Step 2. Thus, this new approach generates two scores on X that are associated with a range of statistically indistinguishable scores on the predicted value of the criterion. Each of these three steps is described next.

Step 1: Establishing Band Width for Y

The first step in the procedure involves the computation of a band width for criterion (i.e., job performance) scores using the equivalent of Equation 1:

$$\text{Band Width} = C \cdot s_y \cdot \left(1 - r_{yy}\right)^{1/2} \cdot \sqrt{2}, \tag{2}$$

where C is the standard score indicating the desired level of confidence for the criterion band, s_y is the standard deviation of the performance measure (e.g., supervisory ratings), r_{yy} is a reliability estimate for the performance measure, and $s_y \cdot \left(1 - r_{yy}\right)^{1/2} \cdot \sqrt{2}$ is the SED between two scores on the criterion measure. There is flexibility in the choice of a value for C, and this choice can be influenced by an organization's willingness to make a Type I or Type II error regarding differences among applicants (Cascio, Zedeck, Goldstein, & Outtz, 1995; see Zedeck, Cascio, Goldstein, & Outtz, 1996, for a more detailed discussion of this issue). We suggest the use of 1.00 as opposed to 1.96 (cf. Cascio et al., 1995) to minimize the number of false positives. Thus, the width of the band on Y will be equal to 1 SED between two scores. In short, if two criterion scores fall within the same band, they are defined as indistinguishable regarding the underlying job performance construct.

Step 2: Establishing the Upper and Lower Limits of the Band on Y

The ultimate goal of this new procedure is to identify those two values of the predictor X that yield predicted values of Y equal to the upper and lower limits of the band on Y (i.e., Y'_{upper} and Y'_{lower}). Therefore, after the width of the band on Y has been computed through Step 1, this band must be superimposed onto some part of the Y scale so that the values corresponding to the upper and lower limits of the band can be established.

The Cascio et al. (1991) banding procedure typically demands the superimposition of bands onto the top of the predictor scale for which the band was computed. In the procedure described here, the extreme top of the Y scale cannot be used because the upper parts of this band would contain values for Y that are too extreme to represent predicted values of Y based on a regression equation relating X and Y scores. For example, suppose we wish to set a band on a selection test that contains scores ranging from 0 to 100, $M = 70$, and $SD = 15$. Suppose further that this test has a correlation of .50 with a criterion, and that this criterion also has a range from

0 to 100, $M = 70$, and $SD = 15$. Suppose finally that, based on these data, we compute a band width for Y of 18 points. If we place this band at the top of the Y scale, then the raw score limits on the band for Y are 82 and 100, and their corresponding Z scores are .80 and 2.00 respectively. Given a correlation of .50, however, the highest possible predicted Z score value of Y is $2 \cdot .50 = 1.00$.

Note that a very high score on one of two positively (but not perfectly) correlated variables will likely be associated with a score on the other variable that is high but not as high as the score on the first variable (i.e., regression to the mean). Thus, there is no value of X that will produce a predicted value of Y equal to the top of the Y scale. Therefore, we recommend the use of the highest possible predicted value of Y as the top of the Y band. This value is computed by first obtaining the unstandardized regression equation predicting Y from X:

$$Y' = a + b \cdot X, \tag{3}$$

where a is the ordinary least-squares (OLS) estimate of the intercept and b is the OLS estimate of the population regression coefficient for predicting Y from X. Equation 3 assumes that personnel specialists have access to the raw data, which is the typical case, or to statistics that allow for the derivation of a and b (e.g., see Footnote 2 for the computation of a and b given the validity coefficient, and means and standard deviations for the predictor and criterion scores).

Next, based on the a and b values obtained using Equation 3, Y'_{upper} is computed using the following regression equation:

$$Y'_{upper} = a + b \cdot \text{Max}\, X, \tag{4}$$

where Max X is the highest observed score on the preemployment test X.

Finally, Y'_{lower} is computed by subtracting the value for the band width obtained in Step 1 from Y'_{upper}.[1]

[1]Prior to implementing the banding procedure, ethnic and gender-based tests for differential prediction should be conducted to examine whether there is homogeneity of slopes across groups (we thank an anonymous reviewer for this suggestion). However, given the notoriously low statistical power of moderated multiple regression, and the fact that the minority group is less numerous than the majority group, it is likely that tests for differential prediction will lead to null findings because of insufficient statistical power, even if the population effect is nonzero (Aguinis, 1995; Aguinis & Pierce, 1998a). Thus, it is suggested that power computations be conducted prior to conducting differential prediction tests using available computer programs (Aguinis & Pierce, 1998b; Aguinis, Pierce, & Stone-Romero, 1994). When statistical power estimates are low (i.e., less than the recommended .80 level), findings regarding homogeneity of slopes across groups should be interpreted with caution.

Step 3: Establishing the Upper and Lower Limits of the Band on X

Once the criterion band has been established, what remains is the identification of a band of X scores that corresponds to the band of indistinguishable scores on Y identified previously in Step 2. To do so, the unstandardized regression equation is used to identify the X scores that produce predicted job performance scores equal to the upper and lower limits of the criterion band. Stated differently, the regression equation is used to identify the X scores that, if entered in the regression equation, would yield predicted values of Y equal to the band limits established in Step 2.

The reason that we use the unstandardized regression equation is that the unstandardized regression coefficient is not affected by some of the most pervasive methodological and statistical artifacts in personnel selection research (i.e., range restriction in the predictor scores and measurement error in the criterion scores; Aguinis & Stone-Romero, 1997; Aguinis & Whitehead, 1997).

Identifying the band limits on X involves simply inserting the relevant information into the regression equation, including the upper and lower values for the band on Y, and solving for X. Consider the following equation:

$$Y'_{upper} = a + b \cdot X'_{upper}. \tag{5}$$

Solving for X'_{upper} yields:

$$X'_{upper} = \left(Y'_{upper} - a\right) / b. \tag{6}$$

Similarly, X'_{lower} is obtained using the following equation:

$$X'_{lower} = \left(Y'_{lower} - a\right) / b. \tag{7}$$

where Y'_{lower} is the lower limit of the band on Y established through Steps 1 and 2.

X'_{upper} and X'_{lower} are the two values for X that produce predicted values of Y equal to the upper and lower limits of the band on actual Y given the regression equation. Of course, for many banding applications, the value of X'_{upper} will simply be the highest value of X. In such cases, Equation 6 is not needed.

Now that the mechanics and logic of the new procedure have been described, the next sections of the article are devoted to (a) illustrating the implementation of this new approach to banding, (b) examining how the resulting band width is affected by variations in test validity and criterion reliability, and (c) comparing band widths produced by the new model to those generated by the Cascio et al. (1991) approach.

ILLUSTRATION

In this section of the article we illustrate how the new procedure can be implemented using results reported by Campion, Pursell, and Brown (1988). Specifically, we focus on the prediction of job performance from scores on a structured interview where interview scores are assumed to be gathered using an interval-level scale. The relevant statistics from Campion et al. (1988) are $a = 114.38$ and $b = 45.41$.[2]

Step 1 requires the computation of the width of the band on Y. This involves the solution of Equation 2. The values for r_{yy} and s_y are .76 and 53.42, respectively, and if we choose a C value of 1.00, then the width of the band on Y is 37.01.

Step 2 calls for the computation of the upper and lower values of the band on Y. The upper value is found by entering the highest available predictor score, in this case 5, as well as the values for a and b in Equation 4. Then, this operation yields a Y'_{upper} value of 341.42. If we then subtract the band width value found in Step 1, we find the Y'_{lower} value, $341.42 - 37.01 = 304.41$.

Finally, Step 3 requires the use of the Y'_{upper} and Y'_{lower} values to find the upper and lower limits of a band of equivalent X scores. If we wish the top of our band on X to be the top of the observed scale, then the upper limit of the band would be the highest observed value of X, 5. The value corresponding to the lower limit is found by inserting the relevant values into Equation 7. This yields an X'_{lower} value of 4.19. Thus, we conclude that all job applicants who score 4.19 or higher on this structured interview are predicted to be equivalent with respect to the variable of ultimate interest, namely job performance.

EFFECTS OF RELIABILITY AND VALIDITY ON BAND WIDTH

Murphy and his colleagues (Murphy, 1994; Murphy, Osten, & Myors, 1995) investigated how variations in test reliability affect the width of the band of equivalence scores using the Cascio et al. (1991) procedure. Likewise, Table 1 shows the values for band width on X obtained using the new procedure for various levels of test validity and criterion reliability. Note that test reliability also affects band width computed using the new procedure. However, this is an indirect effect: Test reliability affects the validity coefficient (or the slope value) that, in turn, affects band width.[3]

[2] In the typical personnel selection situation, the raw data would be available to compute a and b. However, in the present example we did not have access to the raw data. Instead, Campion, Pursell, and Brown (1988) provided the following relevant information: $r_{xx} = .88$, $r_{yy} = .76$, M X for the sample of selected applicants = 4.21, $s_x = .60$, = 278.30, $s_y = 53.42$, and r_{xy} (corrected for range restriction in the predictor and measurement error in the criterion) = .51. Based on this information, we computed Note, however, that Campion et al. did not report for the entire group of applicants, but this mean is almost certain to be lower than the mean of the subset of those applicants who were actually hired. Thus, we assumed for the sake of illustration that it was 1 unrestricted SD unit below the restricted mean, 4.21 – .60 = 3.61. Thus, in computing a we used a value of = 3.61.

TABLE 1
Band Width as a Function of Test Validity (r_{xy}) and Criterion Reliability (r_{yy})

	r_{yy}								
	.95	*.90*	*.85*	*.80*	*.75*	*.70*	*.65*	*.60*	*M*
r_{xy}									
.65	.49	.69	.84	.97	1.09	1.19	1.29	1.38	.99
.60	.53	.75	.91	1.05	1.18	1.29	1.39	1.49	1.07
.55	.57	.81	1.00	1.15	1.29	1.41	1.52	1.63	1.17
.50	.63	.89	1.10	1.26	1.41	1.55	1.67	1.79	1.29
.45	.70	.99	1.22	1.41	1.57	1.72	1.86	1.99	1.43
.40	.79	1.12	1.37	1.58	1.77	1.94	2.09	2.24	1.61
.35	.90	1.28	1.56	1.81	2.02	2.21	2.39	2.56	1.84
.30	1.05	1.49	1.83	2.11	2.36	2.58	2.79	2.98	2.15
.25	1.26	1.79	2.19	2.53	2.83	3.10	3.35	3.58	2.58
.20	1.58	2.24	2.74	3.16	3.54	3.87	4.18	4.47	3.22
M	.85	1.21	1.48	1.70	1.91	2.09	2.25	2.41	

Note. Band width is based on a 5-point test with mean $X = 3.00$ and $s_x = 1.00$, a 5-point criterion scale with mean $Y = 3.50$ and $s_y = 1.00$, and $C = 1.00$.

In Table 1 validity values take on what can be considered typical values ranging from .65 to .20, whereas reliability values for criterion scores range from .95 to .60. The resulting band widths are based on a 5-point test, on an interval-level scale, with $\overline{X} = 3.00$ and $s_x = 1.00$, and a 5-point criterion scale with $\overline{Y} = 3.5$ and $s_y = 1.00$. To facilitate the presentation of results, we kept the value of C constant at 1.00. Also, we chose values of $s_x = s_y = 1.00$ so that $r_{xy} = b$ (see Footnote 2). Consequently, although the new banding procedure uses unstandardized regression equations, we can interpret the results in regard to the more familiar r metric.

Table 1 shows that when criterion reliability is high (.95) and there is a strong relation between the test and job performance (.60), the band width is .53 points (or .53 SDs because $s_x = 1.00$). This band would range from 4.47 to 5.00. If we assume that the distribution of X scores is normal, we can convert the lower limit

[3] In contrast with the Cascio, Outtz, Zedeck, and Goldstein (1991) model, this new approach to banding does not include an explicit correction for measurement error in X scores. The reason for this is that the new procedure focuses on the band on Y. Thus, it includes an explicit correction for measurement error in Y scores (Equation 2). Then, however, there is a need to know which applicants are predicted to be included in this band on Y. Consequently, the regression equations are used to establish the link between predicted Y scores and observed X scores.

(4.47) to a Z score to conclude that this band would include the top 7% of the applicants. Note, however, that if the test in use is difficult, the resulting distribution of scores may be positively skewed and, consequently, the percentage of applicants falling within this band will be smaller. On the other hand, if the test is easy, and the resulting distribution of scores is negatively skewed, the percentage of applicants falling within this band will be larger.

As test validity and criterion reliability decrease, the bands become progressively wider. For example, Table 1 shows that in the presence of poor criterion reliability (.60) and a weak relation between predictor and criterion scores (.20), the band on X is 4.47 points wide. Then, subtracting this width from the highest possible score on X we obtain a band ranging from 0.53 to 5.00. Such a band would include over 90% of the applicants and, obviously, is of little or no use for personnel selection decision makers. This does not mean that the band is somehow inaccurate. Rather, the lack of relation between the test and the criterion produces predicted criterion values that are, for the most part, near the criterion mean. Thus, a large range of test scores is needed to identify a relatively small band of equivalent criterion scores. In such cases, any banding procedure may be of little use. Of course, the same can be said for any preemployment test with a weak relation with the criterion.

Figure 1 includes a graphic representation of the effects of test validity and criterion reliability on band width using the values shown in Table 1. A perusal of Figure 1 and Table 1 indicates that, for the typical range of values found in personnel

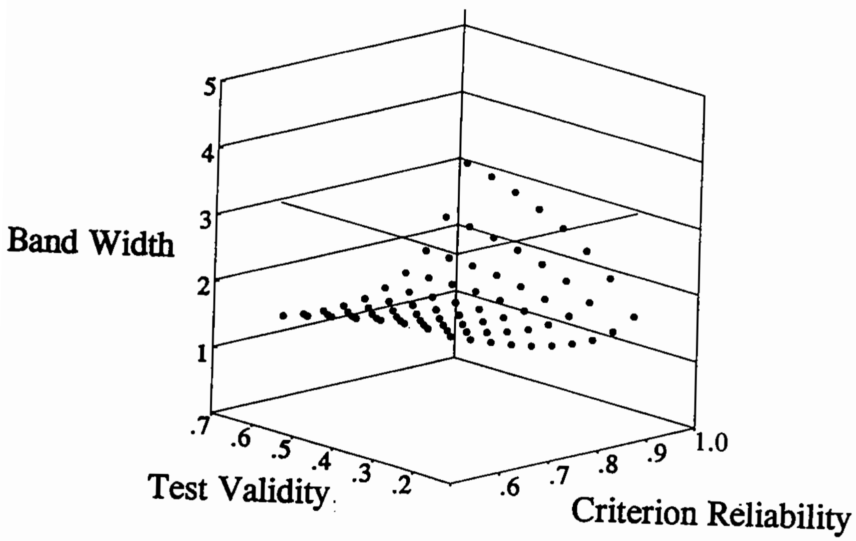

FIGURE 1 Effects of test validity and criterion reliability on band width.

selection, (a) changes in test validity have a greater impact on band width than changes in criterion reliability and (b) the effects of test validity and criterion reliability on band width are not linear. For instance, collapsing across all values for criterion reliability, a .10 increase in test validity from .20 to .30 results in a decrease in band width from 3.22 to 2.15, a decrease of 1.07 points or *SD*s. On the other hand, a .10 increase in the validity coefficient from .40 to .50 results in a decrease in band width from 1.61 to 1.29, a decrease of only .32 points. Similarly, collapsing across values of test validity, a .10 increase in criterion reliability from .60 to .70 yields a decrease in band width from 2.41 to 2.09 points, a decrease of .32 points. Alternatively, a .10 increase in criterion reliability from .80 to .90 yields a decrease in band width from 1.70 to 1.21, or .49 points.

Two conclusions can be drawn from these results regarding efforts to reduce band width and, therefore, increase the degree of discrimination among applicants using the present banding procedure. First, improving test validity has a greater relative impact on band width as compared to improving criterion reliability. Second, when test validity is low, small improvements have a substantial effect in reducing band width. However, equivalent improvements in test validity do not have such an effect in reducing band width when validity is high at the outset.

A COMPARISON WITH THE CASCIO ET AL. (1991) PROCEDURE

An interesting consideration regarding the new approach is how its implementation may lead to different band widths as compared to the procedure currently in use. Once again, we illustrate this point using data from the Campion et al. (1988) study. Earlier, we illustrated that the new approach generated a band ranging from 4.19 to 5.00 (i.e., 1.35 *SD*s). Using the Cascio et al. (1991) approach with the data from the Campion et al. study and a comparable $C = 1.00$ leads to a band width of .29 (i.e., $1.00 \cdot .60 \cdot .35 \cdot 1.41$). This results in a band of equivalent X scores ranging from 4.71 to 5.00 (i.e., .48 *SD*s). The Cascio et al. procedure yielded an ostensibly narrower band. If we convert the lower limits of each band to Z scores, we find that the band computed using the Cascio et al. procedure includes the top 3.36% of applicants, whereas the band computed using the new procedure includes the top 16.60%.

Thus, using the Cascio et al. (1991) procedure leads to the conclusion that applicants whose scores fall between 4.71 and 5.00 are considered to be equivalent regarding the skills and abilities constructs measured by the test. Based on a premise of "acceptable validity," scores falling within this narrower band are presumed to be indistinguishable regarding job performance. In contrast, the new approach uses specific validity information and allows more direct conclusions regarding the extent to which applicants are distinguishable with respect to job performance.

DISCUSSION

The purpose of this article was to describe and illustrate the implementation of a new procedure for the computation of equivalence bands in personnel selection. Whereas the method currently in use takes into account only information about the predictor, the new procedure includes information about the predictor, the criterion, and the relation between the two. Thus, the present approach can be used by personnel specialists inclined to use banding in situations in which criterion data are available. We showed that the new approach for computing band width can be easily implemented. Also, we showed how band widths become progressively wider as test validity and criterion reliability decrease. Finally, we compared the new approach with the one currently in use regarding the resulting band widths.

We conclude that the approach advanced in this article represents an improvement over current methods through its use of criterion information. Specifically, equivalence is determined with respect to job performance and not merely with respect to the construct underlying the preemployment test. Admittedly, criterion information may not always be available. Also, if available, a measure (or a composite of multiple measures) of performance is only an indicator of the performance construct (Cronbach & Meehl, 1955). Thus, it can be argued that any one particular operationalization of performance (or set of operationalizations) that is used to compute the validity coefficient is a less than perfect representation of the performance construct. Nevertheless, using an indicator of the predictor construct (i.e., preemployment tests) alone in computing bands is even further removed from the performance construct than using indicators of the performance construct itself (Cronbach & Meehl, 1955; Nunnally & Bernstein, 1994), which is the variable of ultimate interest in personnel decision making. Consequently, we recommend that the three-step procedure advanced in this article be used in situations in which criterion scores are obtainable.

It should be noted that band widths obtained using the Cascio et al. (1991) procedure are typically narrower than bands produced by the method described herein. The difference in band widths obtained using the two procedures increases as criterion reliability and criterion-related validity values decrease (tables illustrating this phenomenon are available from Herman Aguinis). This should come as no surprise given that the Cascio et al. procedure is unaffected by criterion characteristics. Nevertheless, this difference in band widths has important implications for staffing decision making. If the Cascio et al. procedure is used, bands will be narrower and hiring decisions will, in many cases, be similar to those resulting from a top-down approach. However, decisions regarding the prediction of performance may be less tenable than they could be. Two applicants with indistinguishable scores on X (as determined by the Cascio et al., 1991, procedure) will be assumed to have indistinguishable scores on Y but, due to the less than perfect X–Y relation, there may be little correspondence between bands based on X and actual similarities regarding the performance dimension.

Are Bands Computed Using the New Procedure Inordinately Wide?

An apparent limitation of the present procedure is that it produces inordinately large bands on X when there is a weak relation between the predictor and the criterion, or when there is poor reliability of criterion scores. In the case of a weak relation between X and Y, all of the predicted values of Y will be near the mean of Y. Thus, virtually the entire range of X scores will produce similar predicted scores on Y. However, the fact that the Cascio et al. (1991) procedure yields bands that are narrower in weak X–Y relation situations is potentially misleading. In fact, applicants may be similar regarding the construct underlying X, but may differ substantially with respect to job performance. Consider an extreme example in which $r_{xx} = .80$, $s_x = 5$, and $r_{xy} = 0.00$. The band width based on the Cascio et al. (1991) procedure is 3.1 (i.e., .63 *SD*s) when $C = 1.00$. For the same hypothetical situation, the band computed using the new procedure would encompass the entire range of scores on the predictor. It may seem that the Cascio et al. procedure is, therefore, preferable because the band is narrower and therefore allows distinctions to be made among applicants. However, the new procedure yields a band that allows for a more accurate picture of selection decisions. In fact, it is more accurate to have a band that covers all of the scores on X because the zero-validity predictor offers no reason for believing that any given applicant is more likely to perform better than another. Thus, the relatively small width of the band produced by the Cascio et al. procedure cannot be viewed as advantageous. Instead, it is potentially misleading in that it distinguishes between groups that are equally likely to succeed on the job. Similar arguments can be made with respect to the use of predictors with small (but nonzero) validities.

The aforementioned analytic conclusions are supported by an empirical investigation of the Cascio et al. (1991) model conducted by Siskin (1995). Siskin presented a mathematical model and tables showing the likelihood that the top-ranked person within a band could actually outperform the bottom-ranked person under various conditions of (a) test reliability, (b) criterion sufficiency (the extent to which the criterion measures true performance), and (c) validity. Results showed that, for example, if band width is computed using $C = 2.00$, test reliability = .80, criterion sufficiency = 50%, and the validity coefficient = .70, the difference in expected performance between the top-ranked scorer and the bottom-ranked scorer within the same band is $d = .59$ *SD*s (Siskin, 1995, Table 1, p. 220, line 31). For the same values and a lower validity coefficient of .30, the difference in expected performance between these two applicants drops to $d = .25$ (Siskin, 1995, p. 220, line 35). Thus, as the validity coefficient decreases, there is more uncertainty that high-scoring applicants will outperform lower scoring applicants within the same band. A generalization of these empirical findings from scores within a specific band to the complete range of *X* scores supports our analytic contention that narrow (and seemingly more acceptable) bands obtained using the Cascio et al. procedure in typical conditions of .20–.50 validity are potentially misleading. In actuality, the expected differences in performance levels between applicants may be negligible and, therefore, it is more accurate to represent this situation with (wider) bands computed using validity information.

Closing Remarks

Personnel selection decisions involve predictions regarding the likelihood that applicants will reach specific levels of performance on the job. These predictions regarding expected performance levels are used to make hiring decisions. The model advanced in this article incorporates specific information regarding the predictor–criterion relation. Thus, the new procedure allows personnel selection specialists to make more informed, performance-related predictions regarding applicants' performance that, in turn, allows for more informed selection decision making (cf. Aguinis & Kraiger, 1996). Why should personnel specialists compute band widths assuming some level of acceptable test validity in situations where there is the choice to utilize specific validity information to make more informed decisions regarding applicants' future performance?

ACKNOWLEDGMENTS

The research reported in this article was conducted while Edie Goldberg was at CORE Corporation, Pleasant Hill, CA.

We thank Kevin R. Murphy (Colorado State University), Shelly Zedeck (University of California, Berkeley), and the members of the Behavioral Science Research Group for their comments and suggestions regarding the research reported in this article.

A previous version of this article was presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA, April 1996.

REFERENCES

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21,* 1141–1158.

Aguinis, H., & Kraiger, K. (1996). Intervening in emerging markets: Back to the basics, or go home empty-handed. *The Industrial–Organizational Psychologist, 33*(3), 84–87.

Aguinis, H., & Pierce, C. A. (1998a). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1,* 296–314.

Aguinis, H., & Pierce, C. A. (1998b). Statistical power computations for detecting dichotomous moderator variables with moderated multiple regression. *Educational and Psychological Measurement, 58,* 668–676.

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement, 54,* 690–692.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82,* 192–206.

Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology, 82,* 528–538.

Barrett, G. V., Doverspike, D., & Arthur, W. (1995). The current status of the judicial review of banding: A clarification. *The Industrial–Organizational Psychologist, 33*(2), 39–41.

Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41,* 25–42.

Cascio, W. F., Goldstein, I. L., Outtz, J., & Zedeck, S. (1995). Twenty issues and answers about sliding bands. *Human Performance, 8,* 227–242.

Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4,* 233–264.

Cascio, W. F., Zedeck, S., Goldstein, I. L., & Outtz, J. (1995). Selective science or selective interpretation? *American Psychologist, 50,* 881–882.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Gutman, A., & Christiansen, N. (1997). Further clarification of the judicial status of banding. *The Industrial–Organizational Psychologist, 35*(1), 75–81.

Murphy, K. R. (1994). Potential effects of banding as a function of test reliability. *Personnel Psychology, 47,* 477–495.

Murphy, K. R., & Myors, B. (1995). Evaluating the logical critique of banding. *Human Performance, 8,* 191–201.

Murphy, K. R., Osten, K., & Myors, B. (1995). Modeling the effects of banding in personnel selection. *Personnel Psychology, 48,* 61–84.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49,* 929–954.

Schmidt, F. L. (1991). Why all banding procedures are logically flawed. *Human Performance, 4,* 265–277.

Schmidt, F. L. (1993). Personnel psychology at the cutting edge. In N. Schmitt & W. C. Borman (Eds.), *Personnel psychology in organizations* (pp. 497–515). San Francisco: Jossey Bass.

Schmidt, F. L., & Hunter, J. E. (1995). The fatal internal contradiction in banding: Its statistical rationale is logically inconsistent with its operational procedures. *Human Performance, 8,* 203–214.

Siskin, B. R. (1995). Relation between performance and banding. *Human Performance, 8,* 215–226.

Zedeck, S., Cascio, W. F., Goldstein, I. L., & Outtz, J. (1996). Sliding bands: An alternative to top-down selection. In R. Barrett (Ed.), *Handbook of fair employment strategies* (pp. 222–234). Westport, CT: Quorum.

Zedeck, S., Outtz, J., Cascio, W. F., & Goldstein, I. L. (1991). Why do "testing experts" have such limited vision? *Human Performance, 4,* 297–308.