HUMAN PERFORMANCE, *13*(2), 199–204

A Clarifying Note on Differences Between the W. F. Cascio, J. Outtz, S. Zedeck, and I. L. Goldstein (1991) and H. Aguinis, J. M. Cortina, and E. Goldberg (1998) Banding Procedures

Herman Aguinis
Graduate School of Business Administration
University of Colorado at Denver

Jose M. Cortina
Department of Psychology
George Mason University

Edie Goldberg
Towers Perrin
San Francisco, California

Hanges, Grojean, and Smith (this issue) reaffirmed the Cascio, Outtz, Zedeck, and Goldstein (1991) "traditional" test score banding procedure and argued that the "alternative" method proposed by Aguinis, Cortina, and Goldberg (1998) is problematic. We clarify 4 differences between the traditional and alternative procedures. We suggest once again that the traditional approach be used when evidence regarding criterion-related validity is not available and that the alternative approach be used when this information is available.

Cascio et al. (1991) proposed the use of preemployment test score banding in personnel selection. Banding is an alternative to the strict top-down selection

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

strategy and is based on the premise that pre-employment 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 (e.g., general mental abilities [GMA]). The banding procedure proposed by Cascio et al. uses information regarding test measurement error to compute bands. If two scores fall within the same band, they are considered statistically indistinguishable and secondary criteria (e.g., job experience, ethnicity) may be used in making a hiring decision.

Aguinis, Cortina, and Goldberg (1998) proposed an alternative procedure for the computation of bands. Similar to the Cascio et al. (1991) procedure, the Aguinis et al. approach is based on the premise that pre-employment measures are never perfectly reliable. Unlike the Cascio et al. procedure, the Aguinis et al. approach is based on the additional premise that pre-employment tests are never perfectly valid. Aguinis et al. noted that the Cascio et al. procedure does not explicitly consider the issue of test validity 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, Aguinis et al. also noted that given the imperfect nature of prediction systems in personnel selection, 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 (e.g., GMA) also are indistinguishable or distinguishable regarding the criterion construct (i.e., job performance) may not be tenable. Thus, in addition to information regarding a test's measurement error, the Aguinis et al. approach considers specific evidence regarding the criterion-related validity of the test as well as measurement error of the performance measure used in validating a test.

Hanges et al. (this issue) reaffirmed the Cascio et al. (1991) traditional approach to banding and critiqued the Aguinis et al. (1998) alternative approach. The purpose of this article is to clarify differences between these two approaches to banding and to reaffirm the usefulness of implementing the alternative approach whenever criterion-related validity information is available.

DIFFERENCE NUMBER 1: ASSUMPTIONS REGARDING VALIDITY

Hanges et al. (this issue) argued that the traditional and alternative approaches to banding differ regarding the implicit validation model underlying each of the procedures. Hanges et al. argued that the traditional approach is more consistent with the validation process as it is conceptualized and practiced in personnel selection, whereas the alternative approach is more consistent with classical test theory and the research psychology literature. Hanges et al. contended that the personnel se-

lection validation process is characterized as emphasizing "KSAOs [knowledge, skills, abilities, and other characteristics] because these constructs are believed to be the primary causal agents for individual differences in job performance" (p. 187). They also asserted that although evidence regarding criterion-related validity "can demonstrate that inferences drawn from tests are appropriate, it is more important for the traditional banding procedure that tests are content valid" (p. 188), and content-related evidence can be gathered through a good job analysis. Thus, Hanges et al. maintained that "the lack of attention to criterion information is not problematic for the traditional banding approach" (p. 188). Moreover, Hanges et al. concluded that "the inference that applicants falling within a single traditionally developed band will exhibit equivalent performance on various criteria is reasonable" (p. 188).

We agree that there is a difference in assumptions between the two approaches to banding, and this difference drives differences in the computation of bands. However, we disagree that solely emphasizing KSAOs and ignoring criterion-related information when it is available is in the best interest of personnel psychologists.

It would seem that few others would advocate refusal to consider criterion-related validity information when it is available, as any recent issue of any top industrial psychology journal would attest. In the context of banding, criterion-related information is especially useful because it allows a precise linkage to be made between pre-employment test scores, on which a band must be placed, and criterion scores, which are of ultimate interest in selection situations.

It is certainly not our intention to advocate a dustbowl approach to validation. Validity is best established through multiple means (Landy, 1986), with some of those means being rational (as opposed to empirical). However, we see neither wisdom in ignoring available criterion data nor support for the claim made by Hanges et al. (this issue) that either banding procedure implies the superiority of any particular validation model. We simply suggest that if criterion-related data are available, they should be included in efforts to identify applicants who are likely to perform similarly on the job. The Aguinis et al. (1998) procedure allows such inclusion.

DIFFERENCE NUMBER 2: BAND WIDTH

Hanges et al. (this issue) noted that bands generated using the alternative approach are virtually always wider than bands computed using the traditional approach. This already had been demonstrated empirically by Aguinis et al. (1998). In addition, Hanges et al. provided a derivation and empirical analysis showing that, in virtually all cases, bands produced using the traditional approach are completely subsumed by bands computed using the alternative approach. Moreover, Hanges et al. concluded that, because bands computed using the traditional approach are nar-

rower and included within bands computed using the alternative procedure, scores falling within a traditional band "can frequently be interpreted as being indistinguishable on both the test's latent construct and job performance" (p. 191).

Hanges et al.'s (this issue) analysis provided an eloquent demonstration that applicants falling within the same traditional band also fall typically between the same alternative band. However, Hanges et al. failed to include a discussion of applicants who fall within the wider alternative band and do not fall within the narrower traditional band. Given the data presented by Aguinis et al. (1998) and Hanges et al. showing that traditional bands are often substantially narrower than alternative bands, there should be a large number of applicants classified as indistinguishable regarding predicted job performance using the alternative approach that are screened out using the traditional approach. In the personnel selection literature, these applicants are labeled "false negatives" (i.e., applicants predicted to achieve a sufficient level of performance, yet not selected by the test in question). This exclusion is always problematic. Indeed, it is the very foundation on which banding, as opposed to strict top-down selection, is based. Moreover, given today's U.S. labor market, including an unemployment rate of approximately 4%, organizations cannot afford to screen out potentially successful applicants. Using the traditional approach to banding leads to narrower bands and a reduction of the pool of applicants to be included in a band.

In our view, the fact that the alternative procedure yields wider bands does not decrease "a test's utility for an organization without any advance toward the ultimate goal of a selection system" (Hanges et al., this issue, p. 191). Instead, it gives a clearer picture of that utility. Alternative bands group applicants according to their likelihood of demonstrating similar levels of job performance. Traditional bands group applicants according to their likelihood of demonstrating similar levels of KSAOs. Thus, the alternative procedure is congruent with the ultimate goal of a selection system, which is to make hiring decisions based on predicted performance.

DIFFERENCE NUMBER 3: CHOICE OF CRITERION

Hanges et al. (this issue) noted that a problematic issue regarding the alternative procedure to banding is that practitioners are required to select a criterion measure and use this measure to derive equivalence bands for the performance construct domain. According to Hanges et al., this is a problem because practitioners do not know which criterion should be used to develop test bands. The traditional approach does not rely on criterion information. The alternative approach does. Thus, personnel specialists implementing the alternative approach are faced with a choice of a criterion measure (e.g., supervisory ratings, customer satisfaction ratings).

We acknowledge that criterion information may not always be available. Also, as noted by Hanges et al. (this issue), a measure or a composite of multiple mea-

sures of performance is only an indicator of the performance construct(s). Thus, we agree with Hanges et al. that any one particular operationalization of performance (or set of operationalizations) that is used to compute a validity coefficient is a less-than-perfect representation of the performance construct. Nevertheless, using an indicator of the predictor construct (i.e., pre-employment test) alone in computing bands is even further removed from the performance construct than using indicators of the performance construct itself (Cronbach & Meehl, 1955), which is the variable of ultimate interest in personnel decision making.

DIFFERENCE NUMBER 4: CHOICE OF RELIABILITY COEFFICIENT

An extension of Difference Number 3 described previously is that Hanges et al. (this issue) noted that personnel specialists using the alternative procedure are faced with the need to choose a type of reliability coefficient for the criterion measure. On the other hand, those using the traditional approach do not face this choice because criterion information is not included in the procedure.

We agree with Hanges et al. (this issue) that users of the alternative approach are faced with choosing a specific reliability estimate for the criterion measure. However, we see this presumed difference between the approaches as an actual similarity. Users of the traditional approach need to make a similar choice regarding the pre-employment test in use. In other words, there are numerous sources of errors in measuring, for example, GMA. Thus, personnel specialists need to decide whether, for example, an internal consistency reliability estimate is more appropriate than a test–retest reliability estimate.

Sources of error are always an issue in every measurement system. The traditional approach includes only one measurement (i.e., predictor), whereas the alternative approach includes two (i.e., predictor and criterion). The same choices and challenges faced in assessing measurement error in the predictor become apparent in assessing measurement error in the criterion.

SUMMARY AND CONCLUSIONS

We are very encouraged by Hanges et al.'s (this issue) efforts to more precisely delineate the differences between the traditional and alternative approaches to banding. Our view is that this exchange will help the development and improvement of banding procedures. We hope these developments will allow personnel specialists to appreciate the usefulness of banding as a tool to balance traditional utility and adverse impact considerations in making selection decisions.

This article clarifies four differences between the traditional and alternative approaches to banding. First, the alternative approach to banding explicitly includes evidence regarding criterion-related validity in computing bands, whereas the tra-

ditional approach includes no validity information at all. The assumption of the traditional approach is that specific criterion-related information is not needed as long as a test presents an acceptable level of useful empirical or content validity. The assumption of the alternative approach is that if evidence regarding criterion-related validity information is available, it should be used in computing bands. Second, bands computed using the alternative approach are typically wider than bands computed using the traditional approach. We do not believe that wider bands computed using the alternative approach are less useful. Moreover, these wider bands reflect the fact that predictors used in personnel selection always show less-than-perfect validity. Third, users of the alternative approach are faced with a choice of a criterion measure, whereas users of the traditional approach do not face this choice because the traditional approach does not include criterion information in generating bands. We acknowledge that any criterion measure is an imperfect operationalization of the criterion construct. However, we also believe that including some evidence of criterion-related validity in computing bands generates more accurate information than including no criterion-related validity evidence. Fourth, users of the alternative procedure face the choice of a reliability estimate for the predictor and the criterion. On the other hand, users of the traditional approach only face the choice of a reliability estimate for the predictor. Thus, although users of the alternative approach face the additional challenge of choosing a reliability estimate for the criterion measure, users of both approaches need to make similar choices and face similar challenges regarding the consideration of measurement error and sources of error in computing bands.

In closing, a choice between the traditional and the alternative approach to banding is easy when criterion-related information is not available: The traditional approach is the only alternative. On the other hand, we suggest that the alternative approach be used whenever criterion-related information is available. Why should personnel specialists compute bands without incorporating evidence regarding the criterion-related validity of their pre-employment tests when this information is available?

REFERENCES

Aguinis, H., Cortina, J. M., & Goldberg, E. (1998). A new procedure for computing equivalence bands in personnel selection. *Human Performance, 11,* 351–365.

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.

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

Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41,* 1183–1192.