Test-Score Banding in Human Resource Selection

Technical, Legal, and Societal Issues

EDITED BY HERMAN AGUINIS

2004

CHAPTER 1

Introduction to Test-Score Banding in Human Resource Selection

Herman Aguinis

Consider the following situations:

- John is seeking admission into a graduate school of business in Colorado and obtains a score of 580 points on the graduate management admission test (GMAT). Susan is seeking the same goal and her GMAT score is 570 points.
- Peter, who is African American, takes the Civil Service Examination required of all applicants for the position of police officer in a large city in Texas. Ed, who is White, also takes the same exam. Ed's score is 5 points higher than Peter's on a 100-point scale.
- Lucy, who is Latina, takes an examination required of all applicants for an entry-level firefighter position in a large city in California. Lucy's score is a few points lower than Juan's, who is also Latino.

The traditional approach to selection decision making is to use a strict top-down procedure. That is, selection is made based on who obtained the highest test score. So, in the above vignettes, John would be preferred over Susan, Ed would be preferred over Peter, and Juan would be preferred over Lucy.

Banding is an alternative decision-making procedure that forms groups of applicants whose scores are sufficiently close to each other to render them indistinguishable from one another (Aguinis, Cortina, & Goldberg, 1998, 2000; Cascio, Goldstein, Outtz, & Zedeck, 1995; Cascio, Outtz, Zedeck, & Goldstein, 1991). Assume that the computation of bands indicates that the above pairs of applicants, John-Susan, Peter-Ed, and Lucy-Juan, have indistinguishable scores. If the graduate school of business is seeking to increase the number of female students in the program because

women are severely underrepresented vis-à-vis the relevant population, Susan may be a preferred candidate over John. Similarly, if the police department is attempting to increase the ethnic diversity of its workforce, they may wish to choose Peter over Ed. And, the fire department, who may be trying to increase its gender diversity, may give preference to Lucy over Juan.

Is banding a scientifically legitimate strategy in selection decision making? Does banding lead to a meaningfully greater representation of ethnic and other minorities in the workforce? Are the assumptions underlying the formation of bands of scores tenable? Is the use of banding fair to *all* applicants? Under which conditions is the use of banding legally defensible? Is there a role for policy and social considerations in selection decision making? How can an organization decide whether banding may be an appropriate approach to selection decision making given their specific needs and strategic goals? The chapters included in this book address these and numerous other questions. They provide a comprehensive discussion of all aspects of using banding in selection contexts. Next is a brief description of the key issues addressed by each of the chapters included in this volume.

Cascio, Goldstein, Outtz, and Zedeck provide an overview of demographic changes in the United States and discuss various perspectives in defining who is the best qualified candidate for a position. Then, they discuss the place for policy and social considerations in staffing decisions and provide an overview of the theory, mechanics, and alternative forms of banding. Finally, they provide counterarguments addressing 18 objections raised against the use of banding. These objections revolve around the following non-exhaustive types of issues:

- measurement (e.g., is banding contrary to the linear model?)
- scientific validity (e.g., does banding use an arbitrary grouping of scores and thus create a false impression of scientific basis?)
- logic (e.g., is there a logical problem with banding because it includes a general principle that is only applied in certain cases in not others?)
- statistical issues (e.g., because banding uses a one-tailed test, is it true that the actual confidence intervals for banding are not 95% and 68%, but rather 97.5% and 84%?)
- adverse impact (e.g., is it true that the only purpose of banding is to lessen adverse impact on minorities?)
- legal issues (e.g., is banding a form of race norming?)

Cascio, Goldstein, Outtz, and Zedeck close their chapter by suggesting that experts in assessment for employment purposes provide employers with options for taking, or not taking, social considerations into account when making selection decisions.

Hanges and Gettman provide a detailed discussion of how to implement banding. Specifically, they compare the test-focused approach (e.g., Cascio et al., 1991) and the criterion-focused approach (Aguinis et al., 1998) to forming bands. The comparison of the two approaches to computing bands is expanded and enriched by introducing procedures that originated in the psychophysics literature and by drawing a parallelism between the two banding approaches and two different ways of conceptualizing test fairness. Hanges and Gettman close their chapter by providing guidelines for practice and suggestions for future research.

Guion begins by describing the top-down and banding approaches to selection conceptually. Then, he offers reasons why the topic of banding is so controversial and provides some historical background indicating that the concept of banding is not a particularly new idea. Then, Guion discusses the controversial issue of how to determine the width of a band and provides some practical guidance. Next, he addresses the critical issue of how to make a decision regarding how to choose individuals from within a band and argues in favor of not limiting the discussion to gender and ethnic diversity only; other variables such as educational level, age-related skills or wisdom, variety of experience, and other general attributes deemed to be potentially desirable (or undesirable) can also be used as qualifying (or disqualifying) criteria. Guion closes his chapter by stating that users of banding are faced with three critical decisions: (a) what factors are used in determining band-width, (b) how wide or narrow are the bands, and (c) how to choose among individuals from within a band.

Barrett and Lueke provide a critical review and analysis of court cases involving banding with an emphasis on public service testing for selection purposes. Barrett and Lueke link their review of the legal literature with conceptual issues about banding. For example, they point out that although some court decisions have ruled in favor of banding, these decisions applied to specific circumstances (e.g., a consent decree to remedy past discrimination because banding may reduce adverse impact). Barrett and Lueke argue that banding is contrary to the basic civil service proposition of competitive examination and that the extent of adverse impact reduction by using banding depends on the method used to select individuals from within a band. And, they note that the choice to give preference to minority candidates within a band has been supported by the courts in some cases only and these decisions have been based on specific circumstances. Barrett and Lueke argue that some of the evidence that has been presented in court in favor of banding lacks a scientific basis and they discuss specific examples. Finally, Barrett and Lueke describe the Daubert standards, which guide courts in accepting expert testimony, and contend that they have not been applied to cases involving banding. They use this argument to explain why some courts have rendered a decision in

favor of banding. In closing their chapter, Barrett and Lueke conclude that there is not sufficient legal evidence to support the general use of banding for selection purposes.

Schmitt and Oswald address the question of how much importance is being placed on the construct underlying test scores (e.g., general cognitive ability) and on secondary criteria used in banding (e.g., ethnicity) in the selection decision. In other words, secondary criteria such as ethnicity are given a weight of zero in top-down selection because secondary criteria are not taken into account in making the selection decision (this is true as long as the correlation between test scores and secondary criteria is zero). However, what is the implicit weight given to ethnicity in various types of banding implementation? Schmitt and Oswald provide an answer to this question by using policy-capturing techniques to assess the statistical weights applied to ability tests and group status under eight different strategies of test use including various types of banding techniques. Although they use simulated data, the parameters are based on actual meta-analytic results on ability-performance relationships and subgroup differences. Not surprisingly, results show that the various selection strategies led to varying weights for ability and selection. For example, the top-down selection strategy results in a large weight for ability and a relatively small weight for ethnicity, whereas the top-down within-group selection banding approach yields similar weights for ethnicity and ability. Schmitt and Oswald's simulation demonstrates that the relative weights are not known until the test is given, the selection ratio is known, and the band width is computed. Thus, they issue the warning that using a specific banding technique a priori without knowledge of the selection ratio and band width may not actually yield the intended result (e.g., increased workforce diversity).

Laczo and Sackett report the results of an extensive simulation study investigating expected outcomes of adopting different selection rules. Specifically, Laczo and Sackett's simulation show the effects of using eight selection strategies (including top-down and various forms of banding) on the resulting hiring ratio for minority members and the mean predicted job performance score difference between groups. Also, the simulation includes banding computed on the criterion (i.e., the Aguinis et al., 1998, procedure). First, however, they review results of previous simulation work (e.g., Sackett & Roth, 1991) and draw a number of conclusions for practice. For example, top-down selection produces the highest mean test score among those selected but the difference in mean scores is relatively small. And, as would be expected, banding with minority preference does result in greater minority representation than top-down selection or banding without minority preference. Laczo and Sackett go beyond previous work and simulate a situation where an organization has two pieces of information: bands are formed based on a variable (e.g., cognitive abili-

ties) and selection within a band is done on the basis of a second variable (e.g., conscientiousness). Results show that when no preference is given to minority candidates, banding does not improve the selection outcome in terms of minority hiring or predicted job performance. Also, results show that using criterion-referenced banding yielded a much higher proportion of jobs filled by minority members, but at the expense of lowering predicted job performance scores substantially. Thus, Laczo and Sackett suggest that the usefulness of criterion banding in terms of minority hiring be balanced against lower predicted performance levels.

Schmidt and Hunter argue that banding is internally logically contradictory and thus scientifically unacceptable. In their view, banding violates scientific and intellectual values and, therefore, its potential use presents selection specialists with the choice of embracing the values of science or other important values. Schmidt and Hunter critically examine the historical, political, and social factors underlying the emergence and spread of banding and challenge arguments regarding the benefits of using banding. They conclude their chapter with what they describe as a major test for the field of human resource selection: in their view, intellectual values should prevail over other values and the use of banding should be rejected.

Murphy argues that whether someone supports the use of banding is likely to reflect broader conflicts in interests, values, and assumptions about human resource selection. For example, Murphy notes that attitudes about banding are likely to be affected by whether one's own probability of being selected for a position is increased, decreased, or unaffected by the use of banding as compared to top-down selection. Murphy also argues that one's views about banding are likely to be affected *implicitly* by one's position regarding the trade-off between efficiency and equity. For example, Murphy considers the scenario where one selection strategy leads to a 10 percent increase in productivity and a 2 percent increase in minority hiring, whereas another strategy leads to a 9 percent increase in productivity and a 5 percent increase in minority hiring. Then, he asks the question, "Which strategy is better?" Murphy indicates that questions such as this one have so far been avoided by both opponents and proponents of banding. Murphy closes the chapter by noting the need to develop methods to help organizations answer questions about the difficult comparison and relative importance of efficiency and equity.

Aguinis and Harden propose the use of multi-attribute utility analysis as a tool for deciding whether banding or top-down selection may be a better strategy for a specific organization in a specific context. A multi-attribute utility analysis takes into account not only the potential loss in predictive accuracy of individual job performance, but also key strategic business variables at the group and organizational levels. Aguinis and

Harden describe how a multi-attribute utility analysis allows for the inclusion of key stakeholders in the process, in addition to human resources staff, which is likely to enhance the credibility of the results. Aguinis and Harden provide a step-by-step demonstration of the procedure and argue that, in implementing a multi-attribute utility analysis, selection specialists are able to collect information to decide whether banding should be a part of their selection process and, at the same time, provide evidence that they are not only good technical employees, but also good strategic business players.

In closing, the present volume provides a comprehensive discussion of legal, technical, and societal issues regarding the implementation of banding as a method to interpret test scores in selection contexts. Each of the chapters includes numerous provocative ideas, guidelines for practice, and suggestions for future research.

REFERENCES

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

Aguinis, H., Cortina, J.M., & Goldberg, E. (2000). 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. *Human Performance, 13,* 199–204.

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.

Sackett, P.R., & Roth, L. (1991). A Monte Carlo examination of banding and rank order methods of test score use in personnel selection. *Human Performance, 4,* 279–295.