Test-Score Banding: Objective Option for Social Considerations or Corrosion of Standards and Scientific Practice?

A Review of

Test-Score Banding in Human Resource Selection: Technical, Legal, and Societal Issues, by Herman Aguinis, Westport, CT: Praeger, 2004. 236 pp. ISBN 1-56720-520-8.

Reviewed by
Todd L. Chmielewski
John D. Kraft

Test-Score Banding in Human Resources Selection: Technical, Legal, and Societal Issues, edited by Herman Aguinis, presents a comprehensive analysis of psychometric and human resource findings as well as underlying legal principles that have developed around the use of test-score banding in personnel selection. It also highlights a divisive issue within industrial—organizational psychology.

Traditional testing theory is based on the tenet that individuals who score higher on a test have more ability on a certain trait (i.e., the construct the test is measuring). In test-score banding, individuals with similar scores on a test are grouped together on the basis of the reliability of the selection instruments and their standard errors of measurement. Although methods for determining score bands differ, most share the same general principles. They are derived psychometrically, with the lowest score in the band considered sufficiently similar to the top score. The result is that applicants in the same band are given equal consideration in terms of the test. Although score banding can and has been used for a variety of purposes, the most common is to give organizations more flexibility in terms of hiring a racially and ethnically diverse workforce. The current book examines the question of whether or not test-score banding methods should be utilized via differing perspectives from recognized leaders in the field of industrial—organizational psychology.

In the preface, which also serves as the first chapter, the editor conveys well the controversy of test banding usage and aptly justifies the need for such a book. The remaining chapters can be nicely summarized around three general areas: (a) those who are in favor of test-score banding, (b) the utility of using test-score banding, and (c) those who are against test-score banding. As the title of the book indicates, a large proportion of the chapters delve into the technical, legal, and societal issues involved in score banding in addition to providing guidance on the use of score banding.

In Chapter 2, Cascio, Goldstein, Outtz, and Zedeck begin with an overview of current workforce demographics and consider the implications of social, policy, and technical factors. The authors explore alternative forms of banding and present 18 compelling counterarguments raised by score banding opponents. The chapter is very important in setting the stage for later discussions in the book because most refer to the score banding system originally developed by these authors.

Chapter 3 provides a thorough discussion by Hanges and Gettman on the implementation of score banding. The authors compare the test-focused approach with criterion-based approaches in score banding. They incorporate a psychophysical approach to score banding and couch the chapter in terms of test fairness.

In Chapter 4, Guion describes top-down (i.e., traditional) and banding approaches. He outlines the history of banding, the factors that should be considered when determining band width, how wide or narrow bands should be, and different considerations that could be made to select from individuals within the same band. Guion underscores the idea that score banding is not new.

In Chapter 5, Barrett and Lueke provide a comprehensive review and analysis of the court cases involving banding, specifically emphasizing public service testing. Although the authors acknowledge that there have been cases in which the court has ruled in favor of banding, these decisions applied to specific circumstances (e.g., consent decrees to remedy past disparate impact). They argue that banding is not consistent with the basic civil service premise of competitive examination. They conclude that there is insufficient legal evidence to support the general use of banding. These points are disputed by other authors in this book.

In Chapter 6, Schmitt and Oswald address the issue of weighting (i.e., assigning importance to) both the construct a test is measuring (e.g., job competence) and the secondary characteristic for which banding is being used (e.g., ethnicity). Using simulated data, they analyze various selection processes and conclude that using a specific banding technique a priori will ultimately fail. They conclude that factors such as the selection ratio and bandwidth must be considered before a banding strategy can be successful.

In Chapter 7, Laczo and Sackett report the results of a simulation study that demonstrated the effects of eight different selection strategies on the hiring ratio of minority members as well as the mean predicted job performance scores. As expected, the traditional top-down selection procedures produced the highest score for those selected; however, the difference between mean scores was small. Banding with minority preference did result in greater minority representation. The authors' results show that use of criterion-referenced banding yielded a higher proportion of jobs filled by minority members, but it did so at the cost of job performance. They suggest that a balanced approach may be the best solution.

In Chapter 8, Schmidt and Hunter fervently argue that banding is internally and logically contradictory, hence it is not scientifically acceptable. They trace the history of banding and discuss the social and political factors associated with banding. Their ultimate conclusion is that scientific and intellectual values should prevail over all others. They write that "the wrong choice can result even if all concerned have the best of intentions. Good intentions are not sufficient; this issue requires understanding, thought, and informed choice based on scientific values" (p. 171).

In Chapter 9, Murphy argues that one's view of banding is likely driven by one's implicit view of the trade-off between efficiency and equity. He asks hard questions that he feels both sides of the banding issue have not addressed. He writes the following: "Some of the current controversies over the logic, the acceptability, or the best application of banding are likely to reflect broader conflicts in the interests, values, and assumptions about personnel selection" (p. 177). He stresses the need for better methodologies designed to help organizations make informed decisions regarding efficiency and equity in selection.

In Chapter 10, Aguinis and Harden propose the use of a multiattribute utility analysis as a tool for deciding whether banding or top-down selection may be a better strategy for a particular organization. The authors recognize that different contexts and situations have different needs. Thus, every source of information should be obtained and used in the decision-making process.

Overall, *Test-Score Banding in Human Resource Selection: Technical, Legal, and Societal Issues* is an up-to-date, objective, and definitive treatment of score banding. Most important, it highlights when

score banding may or may not work and can serve as a guide for those organizations that will consider using, or are using, score banding. It provides a comprehensive review of the pros and cons of this selection process, and it points out the key references in the research literature and in the various court challenges that have been made. Test-score banding, under the definition of a test, applies to the use of structured interviews and education and experience ratings as well as written and performance tests. It should be noted that both majority and minority candidates have sued their organizations on the basis of this selection process.

One criticism of these chapters (with the exception of Barrett and Lueke's) is the lack of real-world banding examples. Test-score banding is primarily used by federal, state, and local governments for civil service selection and promotion decisions rather than by private sector employers. This is because most civil service organizations have been required by law to rank order applicants and hire from the top down. Using various rules (e.g., a rule of 3, or 5, or 10), the organization must select first among this grouping and then continue down the lists as the positions are filled. Primarily through the use of consent decrees, some governmental units have been able to use score bands instead of top-down selection. For the federal government, this has changed somewhat. As part of the Homeland Security Act of 2002, Congress authorized federal agencies to use category ratings when there is either a shortage of candidates or a critical hiring need. In the current labor market for federal employment, a large proportion of openings meet this requirement. While giving preference to military service veterans, the appointing official may choose anyone from the top of three quality categories (if there are fewer than three candidates in the top category, the first two may be combined). An appointing official must justify passing over a qualified veteran in the top category to hire a nonveteran. This change is codified in 5 U.S.C. § 3319.

The requirement of these administrative bands states that groupings will consist of individuals with similar levels of job-related competencies, designed to differentiate between the quality of candidates relative to the job. The four requirements stated by the U.S. Office of Personnel Management (the central federal civil service agency) for these category ratings are that they must consider the level of the job; the breadth and scope of the knowledge, skills, and abilities or competencies; the increased level of difficulty or complexity of the job; and the requirements for successful job performance. These determinations should be based on job-analysis findings (5 C.F.R. § 300A). One must wonder if the formal scoring bands, such as those proposed by Cascio et al. and other authors in this book, would result in improvements over more subjective category ratings.

Because the federal government's civil service model is the one used by most state and local governments, it may well be that other public sector organizations will abandon the top-down strict rules of 3, 5, or 10 currently in use. It should also be noted that the authors in this book generally do not criticize the use of administratively determined score bands, such as the use of stanine scores by the military services and other subjectively determined quality rating bands used to cluster similarly qualified individuals for selection and promotion decisions.

Although this book will not change the opinion of anyone who has made up his or her mind on the subject of score banding, it does present an objective view of both sides of the argument. In addition, it can serve as a guide to assist in decision making in the use of banding in human resource selection. This collection of chapters will be useful for both researchers and practitioners.

References

(2002). Homeland Security Act of 2002   Pub. L. No. 107–296, 116 Stat. 2153