Test-Score Banding in Human Resource Selection

Technical, Legal, and Societal Issues

EDITED BY HERMAN AGUINIS

2004

PRAEGER

Westport, Connecticut
London

CHAPTER 10

Will Banding Benefit My Organization? An Application of Multi-attribute Utility Analysis

Herman Aguinis and Erika Harden

The field of human resource selection is faced with a paradoxical situation because using general cognitive abilities and other valid predictors of job performance leads to adverse impact (Aguinis, Henle, & Beaty, 2001; Hough, Oswald, & Ployhart, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt, Sackett, & Ellingson, 2002). Consequently, users of selection instruments are faced with a difficult trade-off: We can choose to use general cognitive abilities tests and risk decreasing the diversity of our organization's workforce, or we can choose to use predictors that will not diminish diversity, but are not as valid as cognitive abilities tests.

Cascio, Outtz, Zedeck, and Goldstein (1991) proposed standard error of the difference (SED) banding as a method to solve the above dilemma. Banding is an alternative to the strict top-down selection strategy that often leads to adverse impact. Banding is based on the premise that 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, then there may be little reason to believe that they differ with respect to job performance (Cascio et al., 1991). In other words, banding groups applicants who have indistinguishable scores. Consequently, job applicants who fall within the same band are considered equally qualified for the job in question. Therefore, choices can then be made among these equivalent applicants based on criteria other than test scores such as diversity considerations (Cascio, Goldstein, Outtz, & Zedeck, 1995).

Aguinis, Cortina, and Goldberg (1998, 2000) proposed a new approach to forming bands that incorporates not only measurement error in the test but also the relationship between test (i.e., predictor) and job performance (i.e., criterion) scores. Aguinis et al. (1998, 2000) argued that the Cascio et al. (1991) model does not explicitly consider the precise predictor-criterion relationship 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 pre-employment tests explain more than 25 percent 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. In short, the Aguinis et al. approach to forming bands of equivalent scores takes into account measurement error in the predictor as well as the predictor-criterion relationship.

Critics of banding have argued that not implementing a strict top-down selection procedure leads to decreased test utility (Schmidt, p. 155, in Campion et al., 2001; Schmidt, 1991; Schmidt & Hunter, 1995). On the other hand, advocates of banding have argued that using banding does not necessarily decrease the utility of a test because using banding can increase an organization's workforce diversity and achieve social goals (Cascio et al., 1995; Zedeck, Cascio, Goldstein, & Outtz, 1996; Zedeck, Outtz, Cascio, & Goldstein, 1991). However, so far, the debate regarding the potential utility loss associated with the use of banding has focused on traditional single-attribute utility analysis. For instance, Sackett and Roth (1991) conducted a simulation to assess the potential loss of economic utility of using banding as opposed to a top-down selection strategy. This simulation adopted a single-attribute utility approach in which the computation of economic utility was based primarily on the size of the predictor-criterion relationship and the subsequent predictive accuracy of individual job performance.

Using single-attribute utility analysis does not answer the key question of whether the use of banding decreases the usefulness of a selection instrument. The use of banding, although it may reduce the predictive accuracy of individual job performance, has the potential to increase the overall utility of a selection instrument because increased diversity has the potential to enhance innovation and creativity, cost reduction in minority recruitment, organizational flexibility, and an organization's public image, among other factors (Cox & Blake, 1991).

The purpose of this chapter is to propose expanding the assessment of the utility of banding from the traditional single-attribute approach to a more context-based multi-attribute perspective. A multi-attribute utility (MAU) 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. Using a MAU analysis provides test users with a more comprehensive decision-making tool to assess whether banding is the right approach for their organization given specific contextual circumstances and objectives. MAU analysis is more comprehensive than the traditional single-attribute utility analysis because it allows for the evaluation of the multiple group and organizational objectives sought by the introduction of banding, as well as the multiple group and organizational consequences of using banding. And, a MAU analysis also allows for the inclusion of key stakeholders in the process, in addition to human resources (HR) staff, which is likely to enhance the credibility of the results (Cabrera & Raju, 2001). Thus, in implementing a multi-attribute utility analysis, selection specialists will be 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 (Ulrich & Beatty, 2001).

The chapter is organized as follows. First, we describe briefly single-attribute utility analysis. Second, we discuss limitations of single-attribute utility analysis and the need to examine broader organizational issues in deciding whether to implement banding. Third, we describe some of the advantages of using multi-attribute, as opposed to single-attribute, utility analysis and provide an overview of various types of systems. Finally, we provide a step-by-step illustration of how to use a multi-attribute utility analysis to gather information in deciding whether banding or a top-down selection approach is most beneficial in a particular organization given specific contextual factors.

SINGLE-ATTRIBUTE UTILITY ANALYSIS MODELS

Brogden introduced the classical single-attribute utility analysis model in 1949. Since then, personnel specialists have used this approach to aid in deciding which selection intervention is likely to lead to the greatest value for the organization. Estimations of utility analysis are often made when HR professionals are considering various courses of action, such as which selection procedure to implement. Prior to Brogden's model, other approaches had been proposed to estimate utility (Taylor & Russell, 1939). However, unlike Brogden's, these earlier models did not allow for the expression of utility in monetary terms. A revision and expansion of Brogden's model by Cronbach and Gleser (1965) included the critical variable of cost of testing. The resulting Brogden-Cronbach-Gleser (BCG) model can be expressed as follows:

$$\Delta U = (N)(T)SD_y r_{xy} \bar{Z}_x - (N)(C), \qquad (1)$$

where ΔU is the increase in average dollar-value payoff resulting from using the selection system in question as opposed to selecting applicants randomly, N is the number of individuals selected using the selection procedure, SD_y is the standard deviation (i.e., amount of variability) of dollar-valued job performance for the job in question, r_{xy} is the correlation coefficient between the selection test scores in question and job performance scores, $\bar{Z}_x$ is the mean standard score on the test for the individuals selected for the job, and C is the average cost of administering the selection system per applicant.

In the past few decades, the BCG model has been expanded and revised to include new procedures for estimating SD_y (e.g., Burke & Frederick, 1984; Cascio & Ramos, 1986) and for integration with capital budgeting models (e.g., Cascio & Morris, 1990) (see Cabrera & Raju, 2001, for a review). For example, Schmidt and Hunter (1983) proposed that SD_y be equated with 40 percent of the average salary for the position in question, and Boudreau (1983a, 1983b) incorporated such factors as variable costs, taxes, and discounting. Nevertheless, in spite of these improvements, the utility model continues to focus on a single central factor: the correlation coefficient between test scores and job performance (i.e., criterion-related validity coefficient).

Although there are published examples of utility analyses demonstrating that very large gains are produced by using valid selection instruments (e.g., Cascio & Ramos, 1986; Schmidt, Hunter, McKenzie, & Muldrow, 1979), HR researchers and practitioners now know that managers may not widely accept or even take into account the large dollar-value utility figures produced (Cronshaw, 1997; Hazer & Highhouse, 1997; Latham & Whyte, 1994; Whyte & Latham, 1997). This realization has put into question the practical usefulness of single-attribute utility analysis.

Next, we describe additional deficiencies of single-attribute utility analysis models and the need for utility analysis models that incorporate not only individual job performance but also other important organizational outcomes.

THE NEED TO GO BEYOND SINGLE-ATTRIBUTE UTILITY ANALYSIS

Single-attribute utility analysis ignores the fact that the impact of a valid selection instrument or procedure goes beyond individual job performance. An improvement in job performance, and the potential subsequent economic gain, is only one of many possible outcomes of changing a selection system (Boudreau 1991; Kaplan & Norton 1996). For example, the introduction of a new selection system can affect an organization's legal exposure, selection ratios for members of protected groups, and an organization's image, among other factors.

A second limitation of single-attribute utility models is that they fail to consider how various organizational constituents are affected by a new selection system. For example, top management, HR, and in-house counsel may have a different appreciation for a system that, in spite of its high degree of psychometric validity, produces adverse impact.

A third deficiency of single-attribute utility models is their failure to consider potential measurement deficiency. In other words, the measure of job performance used in a single-attribute utility analysis may not encompass all aspects of performing a job. Murphy and Shiarella (1997) argued that criterion domains that are multifaceted are better and more realistic for studying the validity of selection tests than the commonly used single-attribute approaches. It is unlikely that a selection instrument, no matter how valid, accurately captures a scope of indicators that cover the entire job performance domain.

The aforementioned deficiencies of single-attribute utility analysis models have led researchers to conclude that there is a need to consider utility models that incorporate individual job performance as well as other outcomes that are also important for organizational success (Edwards & Newman, 1982; Roth, 1994). As noted above, some of the outcomes besides job performance that can be considered include legal exposure, organizational image, and differentiated effects on various organizational constituents, among others. Relevant organizational stakeholders can include HR and labor relations departments, sales and marketing, and upper management.

ADVANTAGES OF MULTI-ATTRIBUTE UTILITY ANALYSIS (MAU)

Multi-attribute utility analysis (MAU) models are decision-making systems that allow for the integration of multiple outcomes to choose among various courses of action (e.g., banding versus a top-down approach to selection). In addition to the advantage of considering multiple outcomes for various organizational divisions or departments, Roth and Bobko (1997) outlined three benefits of using MAU models. First, MAU models are likely to yield a high degree of acceptance and highly credible results. As noted above, recent results on how managers perceive results of single-attribute utility analyses have led to the conclusion that such analyses may lack credibility (Latham & Whyte, 1994). A distinct feature of MAU systems is that managers are an important part of the process and provide input on what outcomes are relevant and will be included in an analysis, how the chosen outcomes will be measured, and the relative weights that will be assigned to each of the outcomes. The high degree of participation and involvement in the process is likely to increase the acceptance and credibility of the results. Second, MAU systems allow for the multidimensionality of performance and, in addition, go beyond individual job performance and can examine the impact of a selection system on performance at the group and even organization level. For example, individual scores can be combined to form a division total performance score, and then combined with all divisions to form an organization-wide score (Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1989). Third, in contrast with single-attribute utility analysis, MAU systems allow for input from various organizational constituents. For example, the HR department may be more concerned with certain outcomes (e.g., an increase in the workforce's ethnic diversity), whereas the marketing department may be more focused on other outcomes (e.g., the ability to recruit an employee with specific knowledge of Spanish media). In short, MAU systems allow for the assessment of different outcomes across organizational units.

Finally, a key advantage of MAU systems is that they can be used to make decisions about the implementation of alternative HR instruments or systems, and results are easier to communicate and more defensible when compared to the results of single-attribute utility analyses (Roth, 1994).

Next, we discuss various types of MAU systems available. Then, we provide a step-by-step description and illustration of how a MAU system can be used to decide whether to implement banding or a top-down approach to selection.

OVERVIEW OF MULTI-ATTRIBUTE UTILITY ANALYSIS MODELS

Several MAU systems have been proposed in economics, management, social sciences, and industrial/organizational psychology (see Roth & Bobko, 1997, for a review of various systems). Some of the systems available include Multiple Attribute Utility Technology (MAUT), Decision Analysis, modified Raju-Burke-Normand (1990), and the Productivity Measurement and Enhancement System (ProMES).

The MAU system that has been most widely used is the MAUT (Roth & Bobko, 1997). Edwards and Newman (1982) proposed this experimental psychology model to aid groups of individuals who hold differing views to overcome their differences and obtain high quality decisions. The theory underlying this model is that a discussion will arise based on a unique model for each group and that this will aid in the group's reconciling its internal differences. In the past it has primarily been utilized in the public sector.

The second MAU system that has a strong research base is labeled Decision Analysis (Roth & Bobko, 1997). This model is based on the ideas put forth by Von Neumann and Morgenstern (1947). The model supports decision makers by providing a standardized process for making a decision (Keeney, 1972; Keeney & Raiffa, 1976). This model is based on a unitary decision maker, as opposed to groups with differing views (Keeney & Raiffa, 1976).

A third MAU model used by Morrow, Jarrett, and Rupinksi (1997) involves a modification of the Raju-Burke-Normand (1990) procedure to evaluate the utility of training interventions. Although the major emphasis of this system is placed on the single attribute of job performance, this system includes such variables as criterion relevance and training transfer and provides a multifaceted utility analysis that could be used for HR interventions in general.

Finally, the ProMES model presents organizations with multisource information on the effectiveness of various organizational interventions (Naylor, Pritchard, & Ilgen, 1980). The ProMES model is based on the assumption that organizational decision makers are able to identify what factors to use to assess organizational effectiveness and the best way to measure each factor. Originally designed to measure productivity in a wide variety of public and private sector organizations, recently it has been established as a MAU method (Roth, 1994).

USING ProMES TO DECIDE ON THE USE OF BANDING

Based on the above description of the various MAU systems available and their potential application to HR interventions (Roth & Bobko, 1997), we believe the ProMES system is a decision-making system well suited to determine whether it may be beneficial to use banding in a specific organizational context. A key attribute of ProMES is its ability to allow organizational decision makers to decide what factors they should use to evaluate the organizational effectiveness of competing interventions and how each of these factors should be measured (see Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1988, for a detailed discussion of ProMES). Next, we offer a detailed description and illustration of the six steps involved in using ProMES to gather information to decide whether banding may be a beneficial approach vis-à-vis a top-down approach.

Step One: Identification of Stakeholders

We start with the assumption that most organizations use a top-down approach. Thus, the first step is to identify those stakeholders who may be affected by the implementation of banding in lieu of the top-down approach currently in use. Possible stakeholders affected by banding include the following: the HR department, the legal department, the sales and marketing department, the public relations department, the manufacturing department, and internal/external customers.

Each department or constituent group would work independently during the MAU process. At the end of the process, each group generate effectiveness scores for banding and top-down selection. Then, these group-level scores would be summated to produce an overall effectiveness score for banding and top-down selection for the organization as a whole.

Step Two: Focus Groups

The second step includes holding focus groups that include 5 to 10 representative members from each stakeholder group. Each focus group would be used to gather information about how the use of banding may affect different outcomes in their unit/stakeholder group vis-à-vis the use of a top-down approach.

Step Three: Identification of Outcomes

The third step involves the identification of outcomes. Outcomes are defined as any effects that the implementation of banding has on each stakeholder group. To identify the outcomes, the focus group facilitator would ask the participants in what ways banding is likely to affect the organization and particularly their unit within the organization. The facilitator would record both positive and negative effects for all participants to view. Then, once an extensive list has been assembled, the group would discuss the list and reduce it down to about six critical outcomes.

Outcomes should be clearly stated. Possible outcomes of implementing banding include the following: traditional single-attribute economic utility, demographic diversity in the organization's workforce, legal exposure, external organizational image, job attitudes, perceptions of fairness, employee tenure and turnover, cost of legal defensibility, cost of implementing banding, organizational climate, societal contributions, and number of grievances.

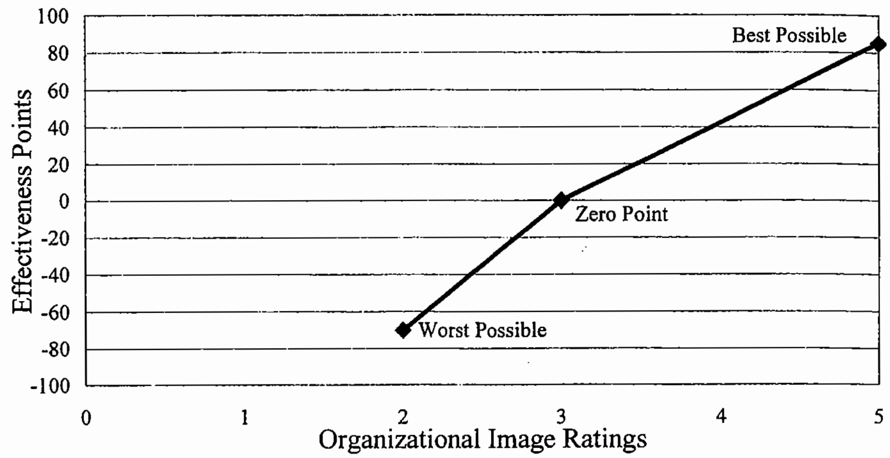
Step Four: Identification of Outcome Indicators

The fourth step entails identifying indicators (i.e., concrete measures) to assess each of the critical outcomes decided on in Step Three. The facilitator would ask participants how they would measure each of the defined outcomes as they relate to the implementation of banding. The participants would start with the first outcome and continue through the list generating ideas for each measure. Typically, each outcome has at least one indicator and may have as many as five. Of course, the greater the number of indicators, the more time and effort would be necessary in the information gathering stage. Ideally, a total of no more than 15 indicators should be used (Pritchard, 1990). Table 1 includes a list of illustrative outcomes and possible indicators. Of course, both the list of outcomes, as well as their indicators, would vary from organization to organization.

Table 1
Examples of Outcomes and Their Indicators

Outcome	Indicators (i.e., measures)
External organizational image	• Survey of applicants during hiring process • Media exposure • Survey of customers
Perceived societal contributions	• Community survey
Legal exposure	• Ratings from legal and HR department staff on expected legal exposure
Cost of legal defensibility	• Expected legal cost (e.g., fees, in-company counsel) associated with expected litigation
Ethnic diversity within the organization	• Selection ratio of members of ethnic minority groups to total number of employees
Effects on number of individuals in protected groups	• Deviation from the 4/5ths rule
Traditional single-attribute economic utility	• Single-attribute utility analysis (e.g., Brogden-Cronbach-Gleser model)
Organizational climate	• Climate survey
Employee tenure	• Archival data
Employee turnover	• Archival data
Number of employee grievances	• Report
Employee job attitudes	• Self-reports (e.g., organizational commitment)
Perceptions of fairness	• Survey of employees • Survey of recent hires

Figure 1
Contingency Graph Showing the Relationship between the Indicator "Organizational Image Ratings" and Effectiveness Points for the Banding Approach to Selection



Step Five: Development of Contingencies

A contingency refers to the relationship between each indicator and overall organizational effectiveness. Contingencies can be displayed in graphs and they show how a change in a specific indicator results in upward or downward changes in organizational effectiveness points. In contingency graphs, the levels of each indicator are shown on the horizontal axis and organizational effectiveness points are shown on the vertical axis. Contingency graphs showing the relationship between indicators and effectiveness points are needed because they allow us to assess the relative impact of changing the value of each indicator on the common metric of organizational effectiveness. Figure 1 illustrates the relationship between one indicator (i.e., organizational image ratings) and effectiveness points for banding. Graphs such as the one in Figure 1 are the result of the implementation of Step Five. The process of creating contingency graphs will become more evident as we progress through our illustration.

Consider the following illustrative situation. Assume a fire department in a large city is trying to decide whether the use of banding would be more beneficial than the use of a top-down selection approach. We are going to use the example of an HR department, but the same process would likely be conducted across other departments. Assume the HR department has identified the following outcomes and indicators:

- Outcome 1: Traditional single-attribute economic utility
  - Indicator: BCG utility estimate
- Outcome 2: Demographic diversity
  - Indicator: Selection ratio for ethnic minorities

- Outcome 3: Organizational image
  - Indicator: Applicant perceptions of organizational image

The focus group would probably develop a larger set of outcomes and indicators. We noted above that a total of approximately 6 outcomes including no more than 15 indicators is a good number. However, selecting the above three outcomes and indicators will suffice to explain the sequence of steps involved in the multi-attribute utility analysis. Once the indicators have been defined, the next step involves identifying the best possible value, the worst possible value, and the zero point for each indicator.

Identification of Best Possible Values

The facilitator would ask the focus group what is the best possible value that could be reached for each of the indicators under ideal conditions for each of the alternative interventions under consideration (i.e., banding versus top-down selection). Ideal conditions differ depending on the indicator in question. For example, the facilitator could ask, "If everything went perfectly, everyone worked as hard as possible in the recruiting effort, and all resources were available to create a very diverse applicant pool, what would be the highest minority ratio we could achieve by using [banding or top-down selection]?" Members of the focus group would provide answers, and initial disagreement is expected. The focus group should continue to exchange ideas until a consensus is reached about the best possible value of each indicator.

In our present example, assume that the focus group discussed best possible values for each of the indicators and reached consensus on values for the banding and top-down systems. Table 2 includes this information.

Identification of Worst Possible Values

Then the facilitator would ask the focus group what is the worst possible value for each of the indicators. To do this, the facilitator could ask focus group members to identify the point of the indicator where major negative consequences would start to occur if the indicator got that bad. For example, this would be the point at which the selection ratio of minorities would lead to unacceptable levels of adverse impact.

In our illustration, assume the worst possible values are identified as shown in Table 3 for the banding and top-down systems.

Identification of Zero Points

Once the best and worst levels have been established, the next step would be to identify the zero point for each indicator. The zero point is the level at which an indicator is neither good nor bad, neither positive nor

Table 2
Best Possible Values for Indicators Using Banding and Top-Down Selection

Indicator	Best Possible Value
Banding	
BCG utility estimate	$7,000
Selection ratio of minorities	30%
Organizational image ratings (1 = poor, 5 = excellent)	5
Top-Down	
BCG utility estimate	$7,500
Selection ratio of minorities	20%
Organizational image ratings (1 = poor, 5 = excellent)	4

Table 3
Worst Possible Values for Indicators Using Banding and Top-Down Selection

Indicator	Worst Possible Value
Banding	
BCG utility estimate	$6,000
Selection ratio of minorities	4%
Organizational image ratings (1 = poor, 5 = excellent)	2
Top-Down	
BCG utility estimate	$6,500
Selection ratio of minorities	2%
Organizational image ratings (1 = poor, 5 = excellent)	1

Table 4
Zero Points for Indicators for Both Banding and Top-Down Selection

	Zero Point
Indicator	Zero Point
BCG utility estimate	$6,500
Selection ratio of minorities	8%
Organizational image ratings (1 = poor, 5 = excellent)	3

negative. The zero point is independent of the use of banding or a top-down strategy for selection. Thus, the zero points for the indicators are the same for both strategies.

To obtain the zero point for each indicator, the facilitator could ask the following question: "What is the point for [each indicator] that if the value were worse, your unit would see negative effects and if the value were better your unit would reap benefits?" As we will see later, the zero points are needed to display contingency graphs in case the relationship between an indicator and effectiveness points is non-linear. Assume the resulting zero points are as summarized in Table 4.

Scaling of Indicators

Once the best possible value, worst possible value, and zero points have been established for each indicator, the next step in generating the contingency graphs is to understand the number of effectiveness points associated with the best and worst possible values. This is done separately for the banding and top-down systems.

To assign effectiveness points to the best possible values for each indicator, the facilitator would ask the focus group participants to rank order the values by the effect they would have on their unit. In other words, the focus group should rank an indicator's best possible value in terms of its overall importance to the unit's work. The facilitator could prompt the group by asking the following question: "If each of the indicators we identified were at the zero point [as shown in Table 4] and only one indicator could be at its established best possible level [as shown in Table 2], which indicator would you choose?" The group eventually reaches a consensus on a rank-ordered list including the best possible value for each of the indicators. In our illustration, assume the focus group decides that the rank order is the same for the banding and top-down systems, and that the order is the following (from most important to least important):

1. Selection ratio of minorities
2. BCG utility estimate
3. Organizational image ratings

The first ranked item (i.e., selection ratio of minorities) is then given the maximum possible amount of points (i.e., +100). The group members are then asked to assign effectiveness points to the other indicators as a percentage of the +100 point maximum. The idea is that the most important indicator is given a value of +100 and the remaining indicators are compared to this one to determine how important each is relative to the most important one.

Assume the group comes to a consensus that the BCG utility estimate is only slightly less important than the selection ratio of minorities. Thus, they assign +90 points to this indicator. The process continues in this same fashion until all best possible values for each of the indicators have been assigned effectiveness points. Assume the resulting effectiveness points are as shown in Table 5. To make the illustration easier to follow, we assumed that effectiveness points were assigned similarly for banding and top-down selection, but this needn't be the case.

Then, a similar process is completed for assigning effectiveness points to the worst possible value for each of the indicators. Indicators are ranked

Table 5
Effectiveness Points for the Best Possible Values for Banding and Top-Down Selection

Indicator	Effectiveness Points
Selection ratio of minorities	+100
BCG utility estimate	+90
Organizational image ratings (1 = poor, 5 = excellent)	+85

according to which would have the most ill effects on the organization if all indicators were at the zero points and one indicator had the worst possible value. The focus group would then proceed to rank the second worst possible value, and so on. The only difference in the process is that the worst possible value seen as most detrimental for the unit is not automatically assigned a score of =100 effectiveness points. To assign a value of =100 to the number one worst possible indicator value would be to assume that this value is equally as bad as the number one best possible indicator value is good. And, this may not be the case. Thus, to assign effectiveness points, the focus group should compare how bad the most harmful worst indicator value is compared to how good the most beneficial best indicator value is. In our example, assume the rank order for the worst possible indicator values yields the following (from the most to the least harmful):

1. BCG utility estimate
2. Selection ratio of minorities
3. Organizational image ratings

Further, assume that the worst possible indicator value for the BCG utility estimate is perceived as 90 percent as bad as the best possible value for the number one best indicator value (i.e., selection ratio for minorities). Given this situation, the BCG utility estimate indicator would be assigned =90 effectiveness points. The other worst possible values are then assigned points as a percentage of the =90 points assigned to the most harmful worst possible indicator value. In sum, the resulting values in our illustration are represented in Table 6.

We can finally create the contingency graphs. Figure 1 displays the contingency graph for the organizational image indicator for banding and

Table 6
Summary of Allocation of Effectiveness Points to Best and Worst Indicator Values for Banding and Top-Down Selection

Indicator	Best Value Effectiveness Points	Worst Value Effectiveness Points
BCG utility estimate	+90	-90
Selection ratio for minorities	+100	-80
Organizational image ratings	+85	-70

Figure 2
Contingency Graph Showing the Relationship between the Indicator "Organizational Image Ratings" and Effectiveness Points for the Top-down Approach to Selection

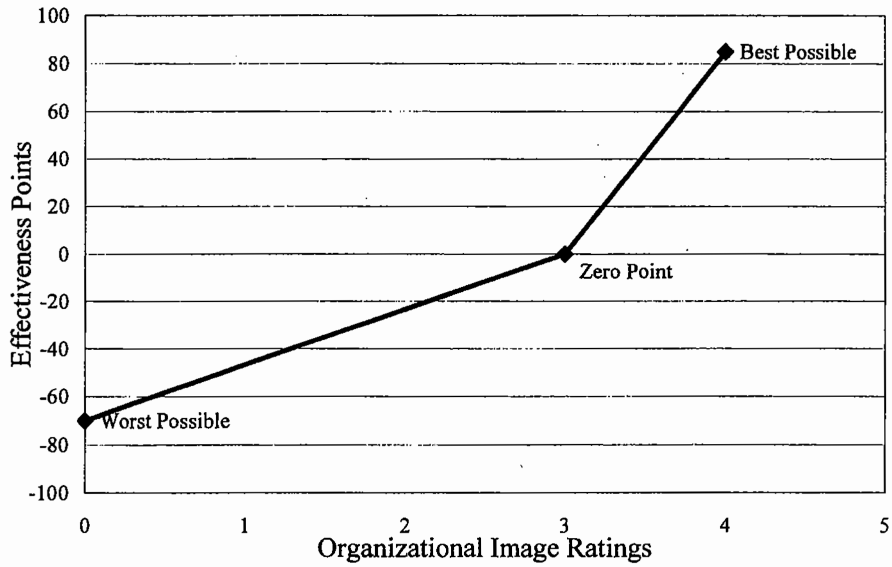
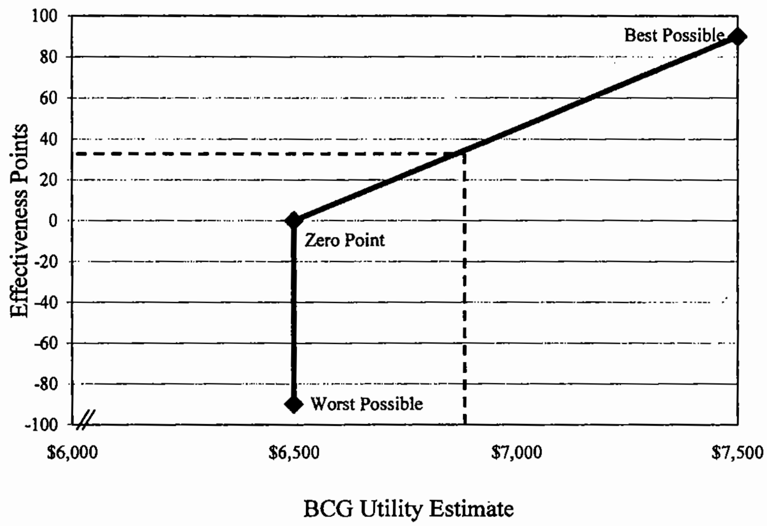


Figure 2 shows the contingency graph for the same indicator for top-down selection. The horizontal axis of each contingency graph ranges from its best possible value to its worst possible value. The vertical axis goes from a high of +100 to a low of =100. To draw the lines in these graphs, we first locate the coordinates and then connect them with lines. Regarding Figure 1, recall that the best possible value for organizational image for banding is 5 (cf. Table 2), and that this value was given 85 effectiveness points (cf. Table 6). Thus, the first coordinate, labeled "Best Possible" in Figure 1, is (5, 85). Also in Figure 1, the "Worst Possible" coordinate is (2, =70), which represents the worst possible value of 2 organizational image points (cf. Table 3) associated with =75 effectiveness points (cf. Table 6). Finally, 3 organizational image points were associated with the zero point regarding effectiveness (cf. Table 4). Thus, the third coordinate is (3, 0). The same process used in drawing Figure 1 was used to draw Figure 2, which shows the contingency graph for the same organizational image effectiveness points relationship for the top-down approach. The same procedure for graphing contingencies is used to display the relationship between each of the indicators and effectiveness points.

Step Six: Making a Decision by Combining Contingencies

To make the decision as to whether banding or a top-down approach may be most beneficial for an organization in a particular situational con-

Figure 3
Contingency Graph Showing the Relationship between the Indicator "BCG Utility Estimate" and Effectiveness Points for the Top-down Approach to Selection

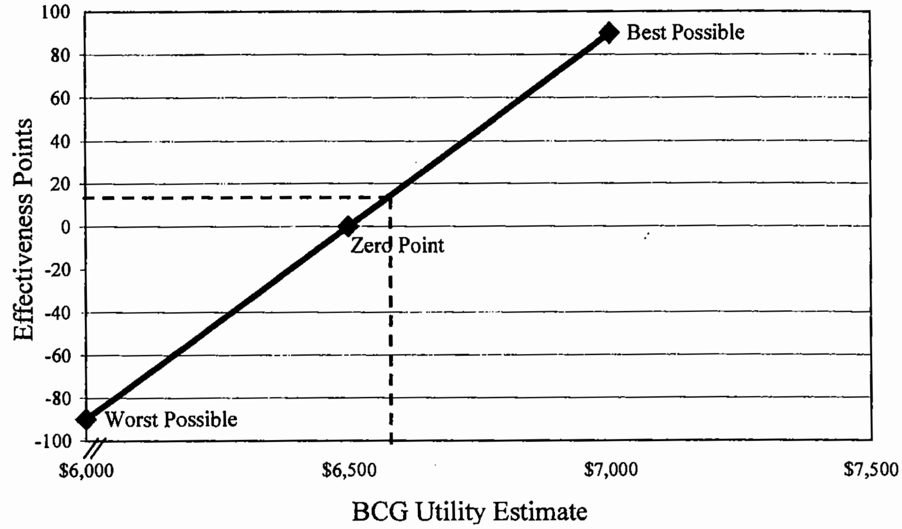


text, we will use the information displayed in the contingency graphs. Specifically, effectiveness points are obtained for banding and top-down selection for each of the indicators. Then, effectiveness points for the three indicators are added and the total effectiveness scores for banding and top-down systems are compared.

Returning to our illustrations, the focus group would consider each indicator separately for the banding and top-down approaches and assign a value that they believe would be most representative if either selection procedure were implemented at the current moment. For example, assume the group starts with the indicator BCG utility estimate for the top-down approach. Thus, the focus group would compute the utility estimate that would reflect the use of a top-down approach. To use a realistic number, we can use data from Cascio et al. (1991), who studied an actual distribution of test scores for 3,377 candidates for jobs as firefighters in large cities in the United States. Cascio et al. (1991) found that strict top-down selection resulted in a utility value of $6,943.99. So, assume that the value obtained by using data from the fire department in our example is $6,900. Figure 3 displays the contingency graph for the relationship between BCG utility estimate and effectiveness points. We draw a line upward from the $6,900 value on the horizontal axis to the slope, and then a line towards the vertical axis. This shows that a value of $6,900 corresponds to 35 effectiveness points.

Now, let's find how many effectiveness points are associated with the potential use of banding also regarding the single-attribute utility analysis indicator. Cascio et al. (1991) reported a utility value of $6,601.23. Similarly, assume that our group estimates the BCG utility value for using

Figure 4
Contingency Graph Showing the Relationship between the Indicator "BCG Utility Estimate" and Effectiveness Points for the Banding Approach to Selection



banding to be $6,600. The focus group would then examine the contingency graph shown in Figure 4 and establish that a utility value of $6,600 corresponds to 17 effectiveness points.

The focus group would then implement a similar process for the remaining indicators. Let's assume that the group establishes that a top-down approach would lead to a selection ratio of minorities of 13 percent. The contingency graph indicates that this number is associated with 32 effectiveness points. And, let's assume that the group expects that banding would lead to a selection ratio of 20 percent. The contingency graph shows that this selection ratio corresponds to 41 effectiveness points. The focus group would then consider the expected organizational image ratings for the use of top-down selection and banding. Let's assume that the focus group decides by consensus that implementing top-down selection today within their fire department would yield an overall organizational image rating of 2. The focus group would examine the contingency graph and determine that this value corresponds to =22 effectiveness points. And, let's assume the group decides that banding would lead to an overall image rating of 4, which the contingency graph shows is associated with 40 effectiveness points.

Table 7 summarizes information resulting from Step 6: the summation of the effectiveness points across indicators. In our illustration, the multiattribute utility analysis resulted in 98 effectiveness points for banding and 45 effectiveness points for top-down selection. Thus, banding is the preferred approach in this particular case. It is interesting to note that

Table 7
Summation of Effectiveness Points across Indicators for the HR Department

Outcome	Indicator	Value		Effectiveness Points	
		Banding	Top-Down	Banding	Top-Down
Single-attribute economic utility	BCG utility estimation	$6,600	$6,900	17	35
Ethnic diversity	Selection ratio of minorities	20%	13%	41	32
Organizational image	Survey of applicants	4	2	40	-22
Total Multi-attribute utility score				98	45

using the traditional single-attribute utility analysis model that only considers individual job performance as an outcome would have led to the opposite conclusion that top-down selection should be the preferred approach.

Finally, as noted earlier in this chapter, the previous example included information gathered from just one group (i.e., HR department staff). In implementing the MAU system, scores originating from the various stakeholder groups are summated to create a grand total organization-level effectiveness points for banding and top-down selection. These grand total scores are used in making the final decision of whether the use of banding may be more beneficial than the use of a traditional top-down approach to selection.

POTENTIAL LIMITATIONS OF USING MAU

Although we are proposing the use of MAU to assess whether implementing banding may be beneficial for an organization, we should acknowledge the following four potential limitations associated with this procedure.

First, what is the decision when the total effectiveness scores for banding and top-down selection are very close? Assume that banding results in 93 points and top-down selection results in 90 points. Is this enough of a meaningful difference to conclude that banding should be the preferred approach? Because the assessments provided by the focus groups are not perfectly reliable, perhaps a three-point difference in favor of banding is not sufficiently large to warrant the implementation of this approach to selection. Following the logic of banding, we suggest that in these situations additional information be taken into account to break the tie. Specifically, additional information can be gathered by reviewing Step 1 and adding one or more additional stakeholder groups. It is likely that the information provided by an additional stakeholder group(s) will break the tie. A second suggestion is related to the second potential limitation of using MAU.

Second, how do we handle disagreements across units? Assume that the HR department's scores indicate that banding should be preferred, whereas scores from the focus group including top management members suggests that top-down selection should be used. As described above, the last step in the MAU process involves adding all scores across stakeholder groups, which in practical terms means that the opinion of each group is given the same weight in computing the grand total effectiveness points for banding and top-down selection. But, does it make sense to give the same weight to the HR department and top management? It could be argued that the use of banding is certainly an important HR issue, but it is also a strategic business issue in which top management may wish to have greater input. Based on the culture of each organization, there is the option of using equal weights across stakeholder groups or assigning differential weights. For example, assuming that there are 5 groups involved, top management's scores may be assigned a weight of 40 percent as opposed to the same weight of 20 percent as the other groups. Nevertheless, the important issue is that the rules regarding how weights will be assigned to the various groups to form the grand total scores be clarified at the beginning of the process. If the scores produced by some groups will be given more weight than the scores produced by others, this needs to be explained clearly together with a rationale for why and how the differential weight system will be applied. Lastly, related to the first potential limitation of MAU noted above, using a differential weight system is also a possible solution for breaking score ties between banding and top-down selection.

Third, Step Five included a description of how to generate contingency graphs showing the relationship between the indicators and effectiveness points. As seen in Figures 1–4, the worst possible, zero point, and best possible coordinates are linked using straight lines. However, it is possible that the relationship between some of the indicators and effectiveness points is not linear. In fact, take Figure 2 showing the relationship between organizational image ratings and effectiveness points for top-down selection. It may be the case that the line linking the worst possible and the zero point coordinates is curvilinear, rather than linear. But, we don't see this as a big problem in the implementation of MAU. First, the indicator-effectiveness functions are generated by using three coordinates (i.e., worst possible, zero point, and best possible values). Errors in specifying a function as linear when in fact it should be nonlinear are minimized when the function is based on three coordinates as opposed to just two (i.e., worst possible and best possible coordinates). Second, the same assumption about linearity is made for the indicator-effectiveness relationship for both banding and top-down selection. So, if curvilinear relationships are reduced to linear relationships, this reduction is made for both approaches and none of the approaches is penalized.

Finally, implementing MAU involves a fair amount of time and effort from a large number of organizational members. In fact, implementing a MAU system may be initially perceived as a daunting task. However, the fact that the process includes a diverse set of outcomes, indicators, and stakeholder groups is precisely what allows the system to be more comprehensive and strategic.

CONCLUSIONS

Banding has been proposed as a method to balance the trade-off between test utility and adverse impact. But, should I implement banding in my organization? Which will be more beneficial given my own organizational context, banding or top-down selection? To answer this question, we need to know what is the utility of using banding as compared to top-down selection. However, thus far, the debate on banding has focused on traditional single-attribute utility only (i.e., test utility). The present chapter suggests that there are numerous organizational goals that should be taken into account when implementing banding. Test utility is just one of them. Using a multi-attribute utility analysis allows organizations to take into account not only the potential loss in predictive accuracy of individual job performance observed by the use of banding, but also key strategic business variables at the group and organizational levels (e.g., organizational image, workforce diversity). Using a multi-attribute utility analysis provides organizations with a more comprehensive decision-making tool to assess whether banding is the right approach for a specific organization given specific contextual circumstances. And, a multi-attribute utility analysis requires the active participation of key stakeholders in the decision-making process, which is likely to enhance the credibility and acceptance of the resulting decision. Implementing a multi-attribute utility analysis allows organizations to quantify the value they place on the various factors involved in choosing between banding and top-down selection. In other words, the various potentially competing values (e.g., test utility versus workforce diversity) are acknowledged and valued explicitly by key organizational stakeholders. And, finally, MAU involves the examination of issues beyond those particular to the HR function. Thus, implementing MAU provides organizational members with a broader picture of how and to what extent this HR intervention is likely to affect the organization at various levels. This is likely to help HR staff provide evidence that they are not only good technical employees but also good strategic business players.

NOTE

We thank Charles A. Pierce (Montana State University) for comments on previous drafts.

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. Outzz, S. Zedeck, and I.L. Goldstein (1991) and H. Aguinis, J.M. Cortina, and E. Goldberg (1998) banding procedures. *Human Performance, 13,* 199–204.

Aguinis, H., Henle, C.A., & Beaty, J.C. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment, 9,* 70–83.

Boudreau, J.W. (1983a). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology, 36,* 551–576.

Boudreau, J.W. (1983b). Effect of employee flows on utility analysis of human resource productivity improvement programs. *Journal of Applied Psychology, 67,* 396–406.

Boudreau, J.W. (1991). Utility analysis for decisions in human resource management. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 621–745). Palo Alto, CA: Consulting Psychologist Press.

Brogden, H.E. (1949). When testing pays off. *Personnel Psychology, 2,* 171–183.

Burke, M.J., & Frederick, J.T. (1984). Two modified procedures for estimating standard deviations in utility analysis. *Journal of Applied Psychology, 69,* 482–489.

Cabrera, E.F., & Raju, N.S. (2001). Utility analysis: Current trends and future directions. *International Journal of Selection and Assessment, 9,* 92–102.

Campion, M.A., Outtz, J.L., Zedeck, S., Schmidt, F.L., Kehoe, J.F., Murphy, K.R., & Guion, R.M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology, 54,* 149–183.

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., & Morris, J.R. (1990). A critical reanalysis of Hunter, Schmidt, and Coggin's (1988) "Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs." *Journal of Applied Psychology, 75,* 410–417.

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., & Ramos, R.A. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology, 71,* 20–28.

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

Cox, T.H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Executive, 5,* 45–56.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions.* Urbana, IL: University of Illinois Press.

Cronshaw, S.F. (1997). Lo! The stimulus speaks: The insider's view on Whyte and Latham's "The Futility of Utility Analysis." *Personnel Psychology, 50,* 611–615.

Edwards, W., & Newman, J.R. (1982). *Multiattribute Evaluation.* Beverly Hills: Sage.

Hazer, J.T., & Highhouse, S. (1997). Factors influencing managers' reactions to utility analysis: Effects of SD_y, method, information frame, and focal intervention. *Journal of Applied Psychology, 82,* 104–112.

Hough, L.M, Oswald, F.L, & Ployhart, R.E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9,* 152–194.

Kaplan, R.S., & Norton, D.P. (1996). Linking the balanced scorecard to strategy. *California Management Review, 39,* 53–79.

Keeney, R.L. (1972). Utility function for multi attributed consequences. *Management Science, 18,* 276–287.

Keeney, R.L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs.* New York: Wiley.

Latham, G., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology, 47,* 31–46.

Morrow, C.C., Jarrett, M.Q.R., & Rupinski, M.T. (1997). An investigation of the effect and economic utility of corporate-wide training. *Personnel Psychology, 50,* 91–119.

Murphy, K.R., & Shiarella, A.H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology, 50,* 823–854.

Naylor, J.C., Pritchard, R.D., & Ilgen, D.R. (1980). *A theory of behavior in organizations.* New York: Academic Press.

Pritchard, R.D. (1990). *Measuring and improving organizational productivity: A practical guide.* New York: Praeger.

Pritchard, R.D., Jones, S.D., Roth, P.L., Stuebing, K.K., & Ekeberg, S.E. (1988). The effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology, 73,* 337–358.

Pritchard, R.D., Jones, S.D., Roth, P.L., Stuebing, K.K., & Ekeberg, S.E. (1989). The evaluation of an integrated approach to measuring organizational productivity. *Personnel Psychology, 42,* 69–115.

Raju, N. S., Burke, M. J., & Normand, J. (1990). A new approach for utility analysis. *Journal of Applied Psychology, 75,* 3–12.

Roth, P. L. (1994). Multi-Attribute utility analysis using the ProMES approach. *Journal of Business and Psychology, 9,* 69–80.

Roth, P. L., & Bobko, P. (1997). A research agenda for multi-attribute utility analysis in human resource management. *Human Resource Management Review, 7,* 341–368.

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.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist, 56,* 302–318.

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., & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology, 68,* 407–414.

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.

Schmidt, F. L., Hunter, J. E., McKenzie, R., & Muldrow, T. (1979). Impact of valid selection procedures on workforce productivity. *Journal of Applied Psychology, 64,* 609–626.

Schmitt, N., Sackett, P. R., & Ellingson, J. E. (2002). No easy solutions to subgroup differences. *American Psychologist, 57,* 305–306.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Tables and Discussion. *Journal of Applied Psychology, 23,* 565–578.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Ulrich, D., & Beatty, D. (2001). From partners to players: Extending the HR playing field. *Human Resource Management Journal, 40,* 293–307.

Whyte, G., & Latham, G. P. (1997). The futility of utility analysis revisited: When even an expert fails. *Personnel Psychology, 50,* 601–610.

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 Books.

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