MODELING THE EFFECTS OF BANDING IN PERSONNEL SELECTION

KEVIN R. MURPHY, KEVIN OSTEN
Colorado State University

BRETT MYORS
University of New South Wales

Selection outcomes under banding are affected by characteristics of the selection system and the applicant pool. This study examined the effects of eight parameters on the proportions hired from higher- and lower-scoring groups: (a) selection ratio; (b) reliability; (b) fixed vs. sliding bands; (d) top-down vs. random within-band selection; (e) preferential vs. nonpreferential selection; (f) mean differences; (g) standard deviation differences; and (h) proportion of applicants from the lower-scoring group. Simulation results were analyzed in a fully-crossed eight-way ANOVA. Higher-order interactions among selection system and applicant pool characteristics had virtually no effect on selection outcomes; the proportion of the applicant pool from the lower-scoring group accounted for nearly half the variance in outcomes. Other important effects are, in order, the effects of standard deviation differences, mean differences, preferential hiring, and the selection ratio. Applicant pool characteristics have considerably more influence on selection outcomes than do selection system characteristics.

The traditional model for personnel selection is to rank applicants on the basis of their observed test scores and select the applicants with the highest scores. This method of selection produces the highest overall selection utility, but it might also produce relatively high levels of adverse impact, particularly for cognitively-demanding selection tests (Gottfredson, 1988; Hartigan & Wigdor, 1989; Sackett & Wilk, 1994). An alternative approach that is receiving increasing attention involves test score banding (Cascio, Outtz, Zedeck & Goldstein, 1991; Sackett & Roth, 1991; Sackett & Wilk, 1994; Scientific Affairs Committee, 1994; for a critical analysis of banding, see Schmidt, 1991). In banding, differences in test scores are sometimes ignored (in particular, scores that are not significantly different from the highest observed score are treated as essentially equivalent), and individuals whose scores fall within the
Figure 1: Applicant Pool, Including Members of Higher-Scoring and Lower-Scoring Groups

The same band are ranked for selection on some basis other than their observed test score (e.g., diversity). Banding represents an extension of procedures that often exist for dealing with tied scores; with banding, some range of scores ends up being treated as if they were tied (Murphy, 1994; Scientific Affairs Committee, 1994). The width of the band is a direct function of the measurement precision of the test (For formulas, see Cascio et al., 1991; Murphy, 1994). Highly reliable tests produce relatively narrow bands, while unreliable tests can produce bands that encompass virtually the entire score distribution (Murphy, 1994).

The most common application of banding is in an effort to reduce the adverse impact of selection tests. Figure 1 illustrates the situation that typically leads to the use of banding. Two groups differ in their test score means, and if individuals are selected solely on the basis of their test scores, relatively few members of the lower-scoring group will receive job offers. However, because all tests are unreliable, one might argue that small differences in test scores should be ignored, and this might lead to you to consider selecting members of both the higher- and lower-scoring groups. In Figure 1, the highest test score is 100 and the test score band is eight points wide, which implies that test scores of 100 to 92 will be treated as essentially identical. This band includes applicants from both groups, and if the individuals in this band are ranked for selection on some basis other than their test scores, it is possible that the employment opportunities of members of the lower-scoring group will increase substantially (Cascio et al., 1991; Sackett & Roth, 1991).
Selection strategies can be designed to achieve a variety of objectives. Strict top-down selection maximizes expected utility (defined in strictly monetary terms), but can produce adverse impact. Alternatives to top-down selection involve tradeoffs, where some decrease in expected financial return is accepted in return for increasing workforce diversity. Until the passage of the Civil Rights Act of 1991 (which expressly forbids this strategy), employers and referral agencies sometimes used score adjustments or within-group norming to reduce adverse impact. Banding appears to achieve the same goal, but is justified in a different way. Advocates of banding (e.g., Cascio et al., 1991) suggest that adverse impact can sometimes be the result of small differences in test scores. Unlike other techniques that group observed scores into broad bands that are treated as functionally equivalent (e.g., stanine scoring, grading systems that translate test scores into letter grades), band widths are determined by the measurement precision of the test. The central assumption of banding is that scores that cannot be reliably distinguished from the highest score should be treated as functionally equivalent, and that individuals with scores in this range should be treated in the same way that applicants whose scores were literally tied would be treated. That is, factors other than their test scores should be used to rank them for selection.

Cascio et al. (1991) showed that banding strategies could be devised that lead to large increases in minority hiring with what they judged to be small decreases in selection utility. Their study focused on one particular strategy that combined sliding bands with top-down, preferential selection among individuals who fell in the same band (each of the components of this banding system are described in detail below). A Monte Carlo study by Sackett and Roth (1991) supported Cascio et al.’s (1991) findings, and illustrated that several decisions about how to implement banding could substantially affect selection outcomes.

**How Selection System and Applicant Pool Characteristics Affect The Outcomes of Banding**

The outcomes of banding depend on characteristics of both the selection system and the applicant pool. The same banding strategy can produce substantially different outcomes in different applicant pools, and very different sets of individuals can be selected from the same applicant pool, depending on the particular details of the banding strategy.

---

1. When applicants are selected on the basis of their observed test scores (i.e., top-down selection), the proportion selected from the higher- and lower-scoring groups is primarily a function of applicant pool characteristics, specifically the differences in means and standard deviations of the two groups. See Sackett and Roth (1991) for analyses of top-down selection in relation to these parameters.
followed. Several recent studies have provided useful insights into the dynamics of banding. Murphy (1994) examined the relationship between test reliability and band width, and showed that unreliable tests could produce extremely wide bands. Sackett and Roth (1994) estimated the proportions hired from higher- and lower-scoring groups and the average test score among those hired for several different banding scenarios, and showed preferential hiring schemes were needed to substantially reduce adverse impact under banding. They also showed that the impact of banding varied systematically as a function of the selection ratio. Cascio et al. (1991) showed that the impact of banding could, in some applications, be similar for bands that varied considerably in size. To date, however, none of the existing studies has considered the full range of parameters that can affect selection under banding.

The purpose of the present study is to examine the separate and joint influence of several selection system and applicant pool characteristics on selection outcomes under banding. We use a Monte Carlo simulation to estimate the impact of each of eight characteristics on selection outcomes, and to estimate the joint impact of various combinations of selection system and applicant pool characteristics on selection under banding. These estimates are useful for two reasons. First, although it is obvious that some selection system and applicant pool characteristics will influence the outcomes of banding (e.g., banding will have a substantial effect on selection from lower-scoring groups if preferential selection is used, or if the differences between groups are small), none of the available studies provide an estimate of the relative importance of the different parameters that affect the operation of banding systems. That is, we know that reliability, preferential selection, selection ratio, and so forth, all have some effect, but it is difficult to determine the absolute or relative importance of each of these factors. Therefore, it is difficult to draw firm conclusions about which aspects of banding are likely to be critical and which are likely to have a small effect in any given context. Second, it is plausible that some factors might interact. For example, the choice between two different banding strategies might make a substantial difference when groups have very different test scores, or when selection ratios are extreme, and might make less of a difference under other circumstances. The literature on banding does not provide a firm basis for making specific predictions about which selection system and/or applicant pool characteristics might interact in affecting selection outcomes, but the possibility of interactions must be seriously considered, especially given the fact that several alternative banding strategies that involve different combinations of factors exist (Cascio et al., 1991; Sackett & Roth, 1991).
Selection System Characteristics

In describing the impact of selection system and applicant pool characteristics on selection outcomes under banding, we will focus on the proportion hired from the higher- and lower-scoring groups. As we will note later, other selection outcomes (e.g., the average test score among those hired, adverse impact) are a direct function of the proportion hired from each of these two groups.

When banding is used in a situation similar to that depicted in Figure 1, the proportion hired from the higher- and lower-scoring groups can be expressed as a function of eight separate parameters, five of which characterize the selection system and three of which characterize the applicant pool (as we note later, several of these parameters also affect hiring outcomes under top-down selection). The selection system parameters that affect the outcomes of banding include the selection ratio, the reliability of the selection test, and three aspects of the banding strategy; that is, whether fixed- or sliding-bands are used, whether there is any systematic preference for members of one group or another when selecting from within the band, and whether within-band selection is done on the basis of test scores.

Selection ratio. The ratio of applicants to available jobs directly affects the organization's ability to be selective in hiring. When only a few applicants are hired, the organization has the luxury of hiring only the top scorers, and selection opportunities for members of the lower-scoring group may be minimal. As the selection ratio (SR) increases, opportunities for members of the lower-scoring group should increase, because the organization is more likely to consider applicants with lower test scores.

Reliability. Murphy (1994) showed that band width can be expressed as a simple function of test reliability (assuming a fixed level of confidence that test scores differ). The less reliable the test, the wider the band. For example, a test with a reliability of .80 will have a 95% test score band 1.23 standard deviations wide. If reliability is .90, the band will be .87 standard deviations wide.

The wider the band, the more likely it is to include applicants from the lower-scoring group. Therefore, all other things being equal, banding should have more impact on selection the lower the reliability of the test used in selection.

Fixed vs. sliding bands. In both fixed and sliding-band systems, the initial test score band is set with reference to the highest observed score (e.g., if the highest score achieved was 94 and the band is 8 points wide, the first band might be set from 94 to 86). In fixed-band systems, individuals are selected from the band until it is depleted. If more jobs remain
to be filled, a new band is established (e.g. from 78 to 86), and individ-
uals are selected from that band until it is depleted, or until all jobs are
filled. In sliding-band systems, a new band is established whenever the
highest-scoring individual is hired or eliminated from the applicant pool
(e.g., if the next- highest score is 92 and the person with a score of 94 is
hired, the band will now extend from 92 to 84).

In general, sliding-band systems provide more selection opportuni-
ties to members of the lower-scoring group, because bands move down
through the distribution more quickly than fixed-band systems, in which
the band must be depleted before it can move.

**Preferential hiring.** One frequent purpose of banding is to reduce the
adverse impact of selection tests. As noted earlier, the use of a selec-
tion system that gives preference to applicants from the lower-scoring
group is critical to attaining that goal (Sackett & Roth, 1991; Sackett
& Wilk, 1994). For example, Cascio et al. (1991) described a diversity-
based banding system in which members of minority groups were given
preference when selecting from within a band. If a test score band in-
cluded 4 minority group members and 10 non-minority applicants, sys-
tems that include preferential selection might require that, when select-
ing from within a band, the 4 minority group members be selected first.

It is important to note that preferential selection applies only to ap-
plicants who fall within the same test score band (i.e., a minority group
member who falls below the band is not given preference to a non-
minority applicant whose score is within the band). The rationale be-
hind preference systems is that when applicants' scores are equal or es-
sentially equivalent, social goals (e.g., increasing workforce diversity) or
secondary criteria might reasonably be given preference over small dif-
fferences in test scores.

**Top-down within-band selection.** The banding system described by
Cascio et al. (1991) included sliding bands, minority preference, and
top-down within-band selection. At first glance, top-down selection of
applicants whose scores are treated as equivalent (because they are in
the same band) might seem contrary to the spirit of banding. If these
test scores are really equivalent, it might not be clear why within-band
selection should be done on the basis of test scores.

If test scores within a band are essentially equivalent, any system of
selecting individuals from within the band might be defensible. The al-
ternative to top-down within-band selection most often cited (e.g., Cas-
cio et al., 1991; Sackett & Roth, 1991) is random selection (with or with-
out minority preference). If any system is appropriate, the criterion for
developing a good system for within-band selection should be the extent
to which it contributes to achieving key goals such as reducing adverse
impact. Cascio et al. (1991) and Sackett and Roth (1991) showed that
a banding system that included sliding bands, minority preference, and top-down within-band selection was highly effective in reducing adverse impact. This particular combination of system features maximizes the probability that members of the lower-scoring group will be considered for employment (because the combination of sliding bands and top-down selection causes bands to move through the distribution quickly), and that they will in fact be hired (because of minority preference).

Applicant Pool Characteristics

Three characteristics of the applicant pool affect the outcomes of banding: (a) the proportion of applicants from the lower-scoring group, (b) group differences in mean test scores, and (c) group differences in the standard deviation of test scores.

Proportion from lower-scoring group. The first applicant pool characteristic that affects the number of individuals from the lower-scoring group actually hired is the proportion of the applicant pool drawn from the lower-scoring group. When the applicant pool contains only a few members of the lower-scoring group, the number actually hired from that group must be low, regardless of the selection strategy. When there are more applicants from this group, the number hired is not so severely constrained, and it is possible to hire larger numbers from that group.

Mean differences. The larger the differences in the mean test scores in the two groups, the lower the likelihood that members of the lower-scoring group will be considered for selection. In Figure 1, the two groups are equally large and have equal standard deviations, which means that the overlap of the distributions of test scores from the two groups is determined solely by the differences in the group means. Under alternate assumptions (e.g., that the standard deviations differ), the mean difference is not the only parameter determining the overlap of the two groups, but it is nevertheless likely to be an important consideration in any comparison of the two score distributions.

Standard deviation differences. The two groups of applicants may differ in both their mean scores and their standard deviations (SD), and these SD differences could substantially affect selection outcomes. If the SD is relatively large in either group, it is possible that the distributions will overlap substantially, even if the means are far apart. It is even possible that the relative position of the two groups will reverse at particular points in the score distribution. For example, if the difference in means is 10 points, and the SDs of the higher-scoring and lower-scoring groups are 10 and 20, respectively, the raw score that corresponds with the 84th percentile in each group will be the same. Because of the larger standard deviation in the lower-scoring group, the scores of the top 16%
of that group will be (on average) higher than the scores of the top 16% of the higher-scoring group.

Reversals in the relative standing of the two groups at different points in the score range will require a combination of small mean differences and large differences in standard deviations. However, even if no such reversals occur, differences in the standard deviations of the two groups could have a substantial impact on selection outcomes because they affect the proportions of the two test score distributions above any given raw score.

Research Question

The eight selection system and applicant pool characteristics outlined above can all affect the outcomes of banding, and various combinations of these characteristics might plausibly lead to different outcomes. In modeling the effects of banding, we simulated all possible combinations of characteristics and examined their effects on selection outcomes via a factorial analysis of variance (ANOVA). This analysis allowed us to determine the effects of each characteristic and each combination of characteristics on the outcomes of selection under banding.

Our rationale for conducting a fully-crossed factorial ANOVA is that existing banding strategies are essentially arranged in a factorial pattern. As Sackett and colleagues have noted (Sackett & Roth, 1991; Sackett & Wilk, 1994), there are a number of banding strategies that can be developed, all of which result from a small number of choices (e.g., to use fixed vs. sliding bands, to use or to avoid preferential selection). As we will show below, all of the banding strategies discussed and analyzed by Cascio et al. (1991) and Sackett and Roth (1991) can be conceptualized as cells in a factorial ANOVA design.

A second analysis examines the relative importance of the simple and combined effects of the five selection system characteristics, assessed over the wide range of applicant pools generated in our simulation study. The rationale for examining selection system characteristics in isolation (i.e., ignoring applicant pool characteristics) is that organizations have more control over the characteristics of their selection systems than over the characteristics of their applicant pools. Although recruitment and outreach efforts may have some impact on applicant pools, the effects of these programs on the types of applicants obtained may be indirect. In contrast, organizations must make specific decisions about how to structure their selection systems, and it is important to know how much impact specific decisions might have on selection outcomes. This second analysis, therefore, provides a closer approximation to the question that must be considered by an organization when deciding how to design its
selection system; that is, how much influence does each selection system characteristic (or combination of characteristics) have on the outcomes of personnel selection with banding.

Method

We used a FORTRAN program developed by Sackett and Roth (1991) to simulate the effects of banding, varying each of the eight factors. The results of these simulations provided the data for our analysis of the effects of the eight selection system and applicant pool characteristics on selection outcomes.

Design

The set of simulations run in this study resulted in an eight-way fully-crossed factorial design, with one observation per cell. The factors in this design were selection ratio (6 levels), reliability (3 levels), fixed vs. sliding bands (2 levels), top-down vs. random selection (2 levels), preference (2 levels), mean differences (3 levels), SD differences (3 levels) and the proportion of applicants from lower-scoring group (3 levels). This fully crossed factorial design yielded 3,888 separate combinations of parameters; the linear model for analyzing this design includes eight main effects, 28 two-way interactions, 70 three-way interactions, and over 130 four-way through seven-way interactions.

Cascio et al. (1991) and Sackett and Roth (1991) described six specific banding systems that are subsumed under our factorial design: (a) fixed band, no preference, random selection, (b) fixed band, minority preference, random selection, (c) fixed band, minority preference, top-down selection, (d) sliding band, no preference, random selection (e) sliding band, minority preference, random selection, (f) sliding band, minority preference, top-down selection. There are two other possibilities, that are also included in our analysis: (g) fixed band, no preference, top-down selection, and (h) sliding band, no preference, top-down selection. Both of these yield the same outcomes as top-down selection (because all within-band selection is done on the basis of test scores), and they would ordinarily not be considered as options in designing band systems. Including these two possibilities, however, allows us to construct a fully-crossed factorial design that in turn allows us to examine the independent effects of each of the eight selection system and applicant pool characteristics, both singly and in combination.
Simulation Parameters

In addition to the substantive factors described above, there are two parameters that can vary in this particular simulation program, sample size and number of replications per simulation run. We chose a large sample size ($N = 600$), to provide stable estimates of the effects of each parameter on selection outcomes and to avoid potential anomalies encountered when the sample is small and the SR is low (e.g. when $N = 50$ and SR = .02, only one applicant is chosen, which makes the proportion from the lower-scoring group either 1.0 or .00). To provide reliable estimates of the effects of banding under each possible combination of parameters, we simulated 500 replications of each scenario, and examined selection outcomes averaging over these 500 replications for each possible combination of the eight parameters varied here. The combination of large sample sizes and a large number of replications of each analysis provided considerable stability and precision in our estimates. For example, in a pre-test run of this simulation, we were able to replicate the results presented in Sackett and Roth (1991) with a level of accuracy that extended to the second decimal point. This suggests that sampling error will have minimal effects on the outcomes of the analyses presented here.

In choosing specific values for each of the substantive parameters in our simulation, we attempted to identify realistic, representative values. The logic for choosing values for each of the parameters in our simulation is laid out below.

Selection ratio. We examined selection outcomes for selection ratios of .02, .05, .10, .15, .30, and .50. We chose to over-sample low selection ratios (i.e., we examined four relatively low SRs and two SRs where there were fewer than four applicants per job), based on the belief that large numbers of applicants are more likely than selection situations where there are only a handful of applicants for each job.

Reliability. We examined selection outcomes for tests with reliabilities of .70, .80, and .90. Murphy's (1994) analysis included a wider range of reliability values, but these figures appear to be reasonably representative of the range of reliabilities reported for cognitive ability tests and pencil-and-paper tests used in personnel selection (Jensen, 1980).

Fixed vs. sliding. We examined both fixed and sliding-band systems.

Top-down vs. random hiring. Within-band selection could be done on either a top-down basis (i.e. hire individual with highest test score first) or on a random basis. Both selection strategies were examined.

Preference. In preferential hiring systems, members of the lower-scoring group are given priority in selection from within the band. In any particular band, members of the higher-scoring group might be eligible for selection only after all members of the lower-scoring group in
that same band have been selected. Following Cascio et al. (1991) and Sackett and Roth (1991), we added the constraint that preferential hiring stops at the point that the selection ratio for members of the lower-scoring group is equal to their proportion in the applicant pool (i.e., we do not allow the hiring rate from this group to exceed the rate that would be obtained under within-group hiring).

Proportion from lower-scoring group. We examined the effects of banding in situations where 10%, 30%, and 50% of all applicants came from the lower-scoring group. These figures seem reasonably representative of the proportion of applicants from racial and ethnic minority groups that might be encountered in a wide range of jobs and regions of the country.

Mean differences. There is an extensive literature examining white-black and white-hispanic differences in mean scores on cognitive ability tests (Gottfredson, 1988; Jensen, 1980). Most studies report differences in the range of .5 to 1.0 standard deviation units; these are the values used by Sackett and Roth (1991) in their simulation study. In our study, we examined the effects of banding when the difference between the two groups was .35, .70, and 1.05 standard deviation units. As we note below, these values represent the approximate 15th, 50th, and 85th percentile of the distributions of group mean differences reported in recent research on group differences in scores on employment-related tests.

Group standard deviations. Sackett and Roth (1991) reviewed research on differential validity to obtain estimates of the relative sizes of the standard deviations for white and black job applicants. Their review, which was limited to studies published prior to 1977, led them to conclude that the standard deviations were usually similar, and that values of .75, 1.0, and 1.25 would approximate the 15th, 50th, and 85th percentiles of distribution of ratios of minority to non-minority standard deviations. Our review of papers published since 1977 led to very similar conclusions. Based on data presented by Cascio et al. (1991), Feild, Bayley, and Bayley (1977), Hennessy and Merrifield (1978), Kesselman and Lopez (1979), Mercer (1984), Morstain (1982), Sandoval (1982), and Zedeck, Outtz, Cascio, and Goldstein (1991), we estimated the mean and standard deviation of the reported SD ratios to be .96 and .24, respectively (estimates of the distribution of group mean differences were obtained from this same set of studies). In our simulations, we used SD ratio values of .70, .95, and 1.20 to approximate the 15th, 50th, and 85th percentiles of the distribution of the ratio of lower-scoring to higher-scoring group standard deviations.
Dependent Variable

In all analyses, the dependent variable was the proportion of positions filled by members of the lower-scoring group. Sackett and Roth (1991) report both this proportion and the average test z-score of those selected; these z-scores provide a concrete indication of the expected utility loss implied by a variety of banding strategies. However, in analyzing the effects of various selection system and applicant pool characteristics on selection outcomes, very little is gained by presenting parallel analyses of both the proportion chosen from the lower-scoring group and the mean z-score among these selected. If the means and standard deviations of both groups are known (as is the case here) and top-down within-group selection is used, the average z-score is a simple transformation of the proportion of positions filled from the lower-scoring group, and analyses of proportion selected and mean z-score are completely redundant. With random within-group selection, the transformation from proportion selected to mean z-score is not exact (the error in approximation increases with band width), but the approximation is still very close, particularly in large samples, and there is still considerable redundancy in analyses of these two potential dependent variables.

Another possible dependent variable proportion of the lower- and higher-scoring groups who are chosen (rather than the proportion of jobs filled from the lower-scoring group). As we will note below, the proportion of each group that is selected is also a simple function of the dependent variable examined here; once the proportions of jobs filled by members of each group are known, it is possible to estimate the average test score among those selected, the selection ratio in each group and the extent to which the selection system produces adverse impact (defined in terms of a comparison of the selection ratios in each of the two groups).

Results

The research design allowed us to estimate all main effects and all interactions (two- through seven-way) except for the eight-way interaction among all of the parameters studied here. This final interaction cannot be estimated because there was a single observation per cell (i.e., here, 2With top-down selection systems, if the proportion selected from the lower-scoring groups is .20, this means that the top 20% of the lower-scoring group will be selected and the remaining jobs will be filled from the top scorers in the higher-scoring group. Because the top scorers in each group are selected, and the means and standard deviations in each group are known, the mean z in from each group can be easily calculated, and the overall mean z is simply a weighted average of the two within-group mean z scores.
TABLE 1
*Main Effects and Nontrivial Two-Way Interactions
From Eight-Way Analysis*

<table>
<thead>
<tr>
<th>Main effects</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection ratio (SR)</td>
<td>.045</td>
</tr>
<tr>
<td>Reliability (R)</td>
<td>.002</td>
</tr>
<tr>
<td>Fixed vs. sliding (F)</td>
<td>.002</td>
</tr>
<tr>
<td>Top-down vs. random (TD)</td>
<td>.001</td>
</tr>
<tr>
<td>Preference (Pref)</td>
<td>.077</td>
</tr>
<tr>
<td>Proportion from low-score group (Prop)</td>
<td>.498</td>
</tr>
<tr>
<td>Mean difference (M)</td>
<td>.089</td>
</tr>
<tr>
<td>$SD$ ratio ($SD$)</td>
<td>.106</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SR $\times$ SD</td>
<td>.029</td>
</tr>
<tr>
<td>Prop $\times$ SR</td>
<td>.018</td>
</tr>
<tr>
<td>Prop $\times$ Pref</td>
<td>.018</td>
</tr>
<tr>
<td>Prop $\times$ M</td>
<td>.021</td>
</tr>
<tr>
<td>Prop $\times$ $SD$</td>
<td>.025</td>
</tr>
</tbody>
</table>

Interactions accounting for more that 0.5% of the variance in selection outcomes are described as nontrivial. Twenty-three of the 28 possible two-way interactions, and all of the higher-order interactions accounted for trivial percentages of variance.

![Figure 2: Selection Ratio $\times$ Standard Deviation Ratio Interaction](image)

*Figure 2: Selection Ratio $\times$ Standard Deviation Ratio Interaction*

an observation represents the mean proportion hired from the lower-scoring group, averaging over 500 replications of a particular banding
scenario). Our first question was whether there was any utility in examining higher-order interactions (i.e., three- through seven-way interactions). We found that all three-way, four-way, five-way, six-way
and seven-way interactions combined accounted for less than 2.2% of the variance in selection outcomes. Furthermore, none of these interactions considered alone accounted for more than 0.5% of the variance. We concluded that it is possible to understand the effects of the eight characteristics considered here on selection outcomes without taking into consideration any of the possible higher-order interactions among applicant pool and selection system characteristics.

Table 1 lists the percentage of variance accounted for by each of the main effects. Two-way interactions accounting for more than 1% of the variance in selection outcomes are also listed. The $SR \times SD$ interaction is illustrated in Figure 2. The interactions between the proportion of lower-scoring applicants and the SR, use of preferential selection, differences in test score means and differences in test standard deviations are shown in Figures 3 through 6.

Table 2 presents the mean proportion of positions filled by members of the lower-scoring group as a function of each of the levels of the eight parameters studied here. This table also shows the proportion of applicants in each group selected, and the ratio of the SRs for the two groups. The 4/5 rule that is often used to define adverse impact suggests that there is a substantial difference in the impact of the selection system on the two groups whenever this ratio goes below .80, which is the case for virtually all of the values shown in Table 2. However, it is important to remember that this table illustrates the effect of each of the eight parameters considered alone. Combining several of these factors in a single selection system can yield dramatically different results than those shown for the simple main effects in Table 2. For example, the sliding band preferential top-down within-group selection method developed by Cascio et al. (1991) yielded essentially equal selection ratios within the higher- and lower-scoring groups across the entire range of overall SRs examined here.

**Breakdown of Variance Due to Selection System Characteristics**

The five selection system characteristics and all two-way, three-way, four-way, and five-way interactions among these characteristics account for 14.6% of the total variance in selection outcomes. As noted earlier, there is considerable utility in assessing the effects of selection system characteristics, independent of applicant pool characteristics. Whereas the characteristics of the applicant pool are not necessarily under the control of the organization, selection system characteristics tend to reflect decisions on the part of the organization, and it is important to understand the implications of decisions organizations make about each of the characteristics of their selection system.
Over 89% of the variance due to selection system characteristics can be attributed to main effects, and 8.6% of this variance can be attributed to two-way interactions; none of the individual two-way interactions accounted for more than 2.5% of the selection system variance. All three-way, four-way, and five-way interactions among selection system characteristics, taken together, accounted for 2.1% of the total variance due
### TABLE 2

**Characteristics of Group Selected as a Function of Eight Simulation Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>% hired from LS</th>
<th>LS group SR</th>
<th>HS group SR</th>
<th>SR ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.02</td>
<td>13.8</td>
<td>.009</td>
<td>.024</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>14.6</td>
<td>.024</td>
<td>.061</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>16.0</td>
<td>.053</td>
<td>.120</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>17.1</td>
<td>.085</td>
<td>.177</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>.30</td>
<td>20.1</td>
<td>.201</td>
<td>.342</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>23.0</td>
<td>.383</td>
<td>.550</td>
<td>.69</td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.70</td>
<td>18.4</td>
<td>.114</td>
<td>.217</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>.80</td>
<td>17.6</td>
<td>.108</td>
<td>.220</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>.90</td>
<td>16.4</td>
<td>.102</td>
<td>.229</td>
<td>.44</td>
</tr>
<tr>
<td>Fixed vs. sliding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed</td>
<td>16.7</td>
<td>.103</td>
<td>.221</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>Sliding</td>
<td>18.7</td>
<td>.116</td>
<td>.216</td>
<td>.53</td>
</tr>
<tr>
<td>Top-down (TD) vs. random</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TD</td>
<td>16.9</td>
<td>.105</td>
<td>.221</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>18.0</td>
<td>.112</td>
<td>.218</td>
<td>.51</td>
</tr>
<tr>
<td>Preferential selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>21.6</td>
<td>.133</td>
<td>.209</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>13.3</td>
<td>.082</td>
<td>.231</td>
<td>.35</td>
</tr>
<tr>
<td>Mean difference^e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.35</td>
<td>23.0</td>
<td>.143</td>
<td>.205</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>17.3</td>
<td>.107</td>
<td>.220</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>12.1</td>
<td>.075</td>
<td>.234</td>
<td>.32</td>
</tr>
<tr>
<td>SD ratio^f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.70</td>
<td>11.4</td>
<td>.070</td>
<td>.236</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td>17.5</td>
<td>.109</td>
<td>.219</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>1.20</td>
<td>23.4</td>
<td>.145</td>
<td>.204</td>
<td>.71</td>
</tr>
<tr>
<td>Proportion from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LS group</td>
<td>5.0</td>
<td>.031</td>
<td>.253</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>.30</td>
<td>16.5</td>
<td>.102</td>
<td>.222</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>30.8</td>
<td>.191</td>
<td>.184</td>
<td>1.03</td>
</tr>
</tbody>
</table>

---

^a Percent of those hired who are members of lower-scoring group.
^b Selection ratio in lower-scoring group.
^c Selection ratio in higher-scoring group.
^d Selection ratio in lower-scoring group/Selection ratio in higher-scoring group.
^e Difference is expressed in $SD^2$ units.
^f Ratio of lower-scoring group $SD$ to higher-scoring group $SD$

### TABLE 3

**Proportion of Selection System Variance Due To Main Effects**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percentage of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection ratio</td>
<td>.313</td>
</tr>
<tr>
<td>Reliability</td>
<td>.020</td>
</tr>
<tr>
<td>Fixed vs. sliding</td>
<td>.018</td>
</tr>
<tr>
<td>Top-down vs. random</td>
<td>.009</td>
</tr>
<tr>
<td>Preference</td>
<td>.529</td>
</tr>
</tbody>
</table>
to selection system characteristics. Proportions of selection system variance accounted for by the main effects of the five characteristics examined here are shown in Table 3.

Table 3 shows that two selection system characteristics have an important impact on the outcomes of banding. Over 50% of the total variance due to selection system characteristics is accounted for by the decision to either employ or avoid systematic preference for members of the lower-scoring groups when ranking applicants whose scores fall within the same band. Over 30% of the total variance due to selection system characteristics is accounted for by the selection ratio. As shown in Table 2, when a large numbers of applications is accepted for each position, the proportion hired from the lower-scoring group decreases. None of the other main effects or interactions has a particularly substantial impact on the outcomes of selection under banding.

Discussion

Two main analyses were carried out, the first examining the main effects and interactions among eight applicant pool and selection system characteristics, and the second examining the effects of selection system characteristics only. The reason for this second analysis is that a detailed assessment of the effects of selection system characteristics provides information about the effects of the organization’s decisions about specific selection strategies on selection outcomes.

Overall Effects of Applicant Pool and Selection System Characteristics

The results illustrated in these Tables 1 and 2 and Figures 2 through 6 suggest that selection outcomes are more strongly affected by applicant pool characteristics than by the particular banding strategy employed. The three strongest effects on selection outcomes were the proportion of lower-scoring group members in the applicant pool, differences in the standard deviations of the two groups, and differences in the mean test score received by each group ($R^2$ values of .498, .106, and .089, respectively). The only selection system characteristics to have an appreciable effect were the use of preferential within-band selection and the overall selection ratio ($R^2$ values of .077 and .045, respectively). The five selection system characteristics (i.e., fixed vs. sliding bands, preferential vs. non-preferential selection, top-down vs. random selection, test reliability, and selection ratio), and all of the two-, three-, four-, and five-way interactions among these characteristics, taken together, explain 14.6% of the variance in selection outcomes.
The interaction effects shown in Figures 2 through 6 suggest that some combinations of selection system characteristics and applicant pool characteristics make a difference. In particular, the effects of the selection ratio on the proportion selected from the lower-scoring group are strongest when the SD in that group is smaller than the SD in the higher-scoring group (See Figure 2). When applicants from the lower-scoring group are grouped closely around their group mean, there may be very few members of this group with scores high enough to be considered for selection unless the selection ratio becomes large.

The effects of the selection ratio are progressively stronger as the proportion of applicants from the lower-scoring group increases (See Figure 3). The same logic holds here as in the paragraph above. When there are only a few individuals chosen, the proportion of applicants from the lower-scoring group may not make much of a difference, because few members of this group will ever be considered for selection. Similarly, the effects of preferential selection are relatively small when the proportion of applicants from the lower-scoring group was small (See Figure 4). With few applicants eligible for preferential selection, the decision to use or avoid this practice may make relatively little difference. Similarly, the effects of mean differences and standard deviation differences between groups become stronger when there are many applicants from the lower-scoring group than when there are few (See Figures 5 and 6). With relatively few applicants from this group, the proportion of positions filled from this group may remain low even if the difference in group means is relatively small or if the SD of the lower-scoring group is relatively large.

The proportion of applicants from the lower-scoring group has such a strong impact on selection outcomes that it is possible this factor masks the effects of other important variables. To determine whether our results were unduly affected by this factor, we repeated the analysis for the seven remaining factors, treating proportion from the lower-scoring group as a replications factor. That is, we carried out three separate seven-way ANOVAs for the simulations carried out on applicant pools with 10%, 30%, and 50% from the lower-scoring group, respectively. The results of these analyses closely paralleled those shown in Table 1. In all three analyses, all higher-order interactions combined explained trivial amounts of variance, and in all three analyses, the same main effects were important as in the analysis shown in Table 1. That is, in all three seven-way ANOVAs, the important determinants of selection outcomes were (in order) differences in group standard deviations, mean differences, the use of preferential hiring, and the selection ratio. The few two-way interactions that accounted for non-trivial variance (i.e., more than 1% of the variance) were also similar to those shown in Table 1 and
Figures 2-6. On the whole, these analyses suggest that the effects of selection system and applicant pool characteristics on selection outcomes under banding are relatively simple, and that many of the decisions faced when choosing a banding strategy (e.g., fixed vs. sliding bands, top-down vs. random selection) have only a small impact on the proportion of jobs filled from the lower-scoring group.

Many of the findings reported here could probably be anticipated on the basis of a simple logical analysis of selection systems. It is no surprise that the proportions of each group in the applicant pool has an important effect on the proportions hired, that preferential hiring increases selection rates from the lower-scoring group, or that differences in the means and standard deviations in the two groups are important. What may be surprising is the relative magnitude of the effects. For example, Murphy (1994) showed that bands can include very large portions of the test score distribution, even when tests exhibit acceptable levels of reliability. The present analysis suggests that, in the context of the other factors that affect selection under banding, the effects of variation in test reliability may be quite small (i.e., $r^2 = .002$). Similarly, previous studies suggest that the decision to use fixed vs. sliding bands or top-down vs. random within-band selection could be important, but the present analysis suggests that the overall influence of these decisions is likely to be small (i.e., $r^2 = .002$ and $r^2 = .001$, respectively), at least in comparison with the influence of applicant pool characteristics.

**Effects of Selection System Characteristics**

Selection system characteristics are all to some degree under the control of the organization, and therefore deserve special scrutiny. Organizations must make decisions about the use of fixed vs. sliding bands, preferential vs. non-preferential selection, and top-down vs. random within-band selection. Reliability is not as easy to directly control, although organizations often have the opportunity to choose among several possible tests. The selection ratio is probably hardest to directly control, although as we will note later, even this factor often reflects the organization's decision to hire as soon as a good applicant becomes available rather than to wait until enough applications have been received to yield a low selection ratio.

Table 3 shows that a very substantial portion of the variance due to selection system characteristics can be accounted for by the simple main effects of preferential vs. non-preferential hiring and the selection ratio. To some extent this finding is redundant with those of the analyses discussed above; these two factors were the only selection system characteristics to have an appreciable effect on selection outcomes. What
was not necessarily so apparent from that analysis was how little influence other selection system characteristics, considered either singly or in combination, had on selection outcomes. For all practical purposes, the decision to use fixed vs. sliding bands, to use tests with low, moderate, or high reliability (i.e., reliabilities of .70, .80, or .90), or to use top-down vs. random within-band selection had little effect on selection outcomes. Analyses presented by Sackett and Roth (1991) and by Murphy (1994) suggest that these factors can, in some specific applications, make a difference. However, in general, decisions about these specific factors may have little impact on the outcomes of selection under banding.

The fact that there are few meaningful interactions between selection system and applicant pool characteristics is important because it implies that the results shown in Table 3 generalize across applicant pools. That is, for any specific organization faced with decisions about how to configure its selection systems, two specific decisions are likely to be very important (i.e., whether to use preferential selection and whether to consider a large number or a small number of applicants). Other decisions (e.g., whether to use fixed or sliding bands, whether to use top-down or random selection, whether to use tests with reliabilities of .9, .8, or .7) have relatively small effects on selection outcomes.

Limitations and Implications

One caution must be observed when comparing the effects of various selection system and applicant pool characteristics on selection outcomes under banding. Every effort has been made to choose levels for each factor in the eight-factor model that are realistic, or that reflect values of these parameters reported in the research literature, but very different conclusions might be reached by other researchers who choose different values for key parameters. For example, the importance of the proportion of lower-scoring group members in the applicant pool would be substantially smaller if we had simulated banding with proportions of .10, .15, and .20 rather than .10, .30, and .50. Nevertheless, the results of this study provide considerable insight into the variables that affect selection outcomes under banding, and they suggest that these variables have generally straightforward and predictable effects.

Perhaps the most important message of Table 1 is that there are no important higher-order interactions. That is, the effects of applicant pool and selection system characteristics on selection outcomes in banding are relatively simple, and can be understood in terms of a small number of parameters. Table 1 lists five main effects and five two-way interactions that, taken together, account for more than 94% of the variance in selection outcomes. The five main effects alone account for more that
80% of the variance in selection outcomes. Thus, although the potential effects of the eight selection system and applicant pool characteristics studied here on banding outcomes are extremely complex (it is possible to statistically estimate over 250 separate independent sources of variation in selection outcomes), the actual effects are relatively simple. In comparison to the effects of applicant pool characteristics, the effects of many possible variations on banding strategies may be quite small. Similarly, the effects of the five selection system characteristics, considered alone, appear to be relatively simple. Over 80% of the total variance due to selection system characteristics is accounted for by two simple main effects, and the net effects of all interaction terms are virtually trivial.

The analyses presented here suggest that the best single strategy for increasing the proportion hired from the lower-scoring group is to change the applicant pool (e.g., through recruiting) rather than to modify the selection system after the fact. Banding represents an after-the-fact strategy for minimizing the effects of group difference, and it can be highly effective (Cascio et al., 1991). On the whole, however, organizations are more likely to minimize adverse impact (and less likely to see their selection systems challenged in court) if they can reduce adverse impact before the fact, by recruiting groups of applicants who are more similar. In particular, Table 2 suggests that recruiting the best applicants in the lower-scoring group might have substantial payoff, because it could raise both the mean and the standard deviation of that group (both of which are likely to increase representation from the lower-scoring group). Whether organizations interested in reducing adverse impact can successfully attract and hire the best applicants from lower-scoring groups remains to be seen; competition for such applicants is likely to be intense. However, if it is possible to change the applicant pool, this is more likely to have an impact on adverse impact than changes to the selection system.

In situations where it is not practical to substantially change the applicant pool, the best strategy for minimizing adverse impact is clearly to take as few applications as possible (thus raising the SR) and establish an explicit preference for members of the lower-scoring group when selecting from within the band. This, of course, assumes that one can increase or decrease the size of the applicant pool without substantially changing its characteristics (e.g., the mean, SD, difference between groups, etc.). In some instances, this might be feasible. Organizations might either make decisions as soon as a few applicants become available (yielding a high SR), or defer their decisions until a large number of applications have been received (yielding a low SR). The results presented here suggest that adverse impact is most easily addressed with banding when relatively few applicants per job are considered. Note, however, that failure
to consider a large number of applicants, in a situation where it might be feasible to defer a decision until the SR is low, will substantially reduce the economic utility of testing.

Finally, it is important to note that banding is controversial. For example, Schmidt (1991) criticized banding on both practical and logical grounds (e.g., one practical objection to banding is that it leads to a potentially large reduction in expected utility. See Murphy and Myors, in press, for a detailed analysis of Schmidt's logical critique of banding). Recent papers by Gottfredson (1994) and Sackett and Wilk (1994) show that banding may lead one to overemphasize some types of decision errors (e.g., false negative decision errors in selecting from lower-scoring groups) and ignore others (e.g., false negative errors in selecting from higher-scoring groups), and that banding can lead to outcomes identical to those obtained under now-illegal test score adjustment strategies. The analyses presented here do not resolve the question of whether banding is appropriate or desirable (See Schmidt, 1991, and Scientific Affairs Committee, 1994, for comments on the acceptability of banding); the ultimate resolution of the question of whether banding is advisable is likely to depend as much on the decision-maker's values as on the empirical research on banding (Sackett & Wilk, 1994). Thus, while the analyses presented here do not answer the question of whether one should or should not pursue banding as a selection strategy, they do shed some light on the factors that are likely to have a strong or a weak impact on the characteristics of the individuals who are selected using banding.

REFERENCES


