Multiple Dimensions of Procedural Justice: Longitudinal Effects on Selection System Fairness and Test-Taking Self-Efficacy

Donald M. Truxillo*, Talya N. Bauer and Rudolph J. Sanchez

To demonstrate the multidimensionality of test fairness, we examined the reactions of 246 police applicants to two consecutive selection tests (written and video-based) in terms of eight dimensions of fairness. As hypothesized, each test was seen as more fair in terms of certain dimensions. Furthermore, test fairness measured immediately after each test predicted perceptions of overall selection system fairness measured after candidates received their test results and after controlling for applicants’ selection outcomes (i.e., whether they were eligible for further consideration in the selection process). Job-relatedness/content for the video-based test interacted with test score to affect test-taking self-efficacy. Our discussion focuses on the multidimensionality of test fairness, the contribution of these dimensions to overall selection system fairness, and the consideration of these dimensions in selection system design.

The importance of applicant reactions to selection system characteristics has become increasingly recognized. Such reactions have been found to affect numerous outcomes such as test-taking motivation (Chan et al. 1997) and organizational attractiveness (Bauer et al. 1998; Macan et al. 1994; Smithe et al. 1993; Truxillo and Bauer 1999). Moreover, negative applicant perceptions have been proposed to relate to subsequent legal action (Seymour 1988). Gilliland’s (1993) model of reactions to selection systems has guided much of the research in this area (e.g., Bauer et al. 1998; Ployhart and Ryan 1998). Grounded in organizational justice theory (e.g., Greenberg 1990; Lind and Tyler 1988), the model proposes that selection system characteristics are considered fair to the extent that they comply with or violate certain procedural and distributive justice rules. The extent to which applicants believe the selection system complies with or violates these rules leads to overall fairness perceptions, which in turn lead to individual and organizational outcomes.

However, several gaps still exist in the applicant reactions literature. First, with few exceptions (e.g., Bauer et al. 1998), research has tended to use global measures of fairness when comparing selection procedures. Such an approach ignores the complexity and multidimensionality underlying applicant reactions to selection systems as described by applicant reactions models (e.g., Arvey and Sackett 1993; Gilliland 1993). Moreover, the contribution of the fairness of individual tests to perceptions of overall selection system fairness has not been explored. Finally, the interaction effects of perceived job-relatedness and test performance on specific (e.g., multiple-choice) test-taking self-efficacy have not been examined.

The present field study filled these gaps in the literature in several ways. First, by using multidimensional measures of fairness, we explored the concept that the perceived fairness of tests is complex and that a given test format can be perceived as fair in some ways but not in others. We also tested Gilliland’s (1993) proposal that these individual dimensions of fairness will affect the overall perceived fairness of the selection system. Specifically, we explored the relationship between test fairness (i.e., the individual dimensions) measured at the time of testing and the overall fairness of the selection system after applicants have received their results. Finally, we extended previous work (e.g., Bauer et al. 1998) by investigating the combined effects of job-relatedness perceptions and test performance on specific (e.g., multiple-choice and video-based) test-taking self-efficacy. We explored these issues using a longitudinal design over a four-month long selection process for police recruit.

A true comparison of multiple-choice (MC) and video-based (VB) formats was not possible in this field study because we could not control
for content and order effects in this field setting. Rather, the purpose of this study was to show the importance of considering multiple dimensions of test fairness by demonstrating these two formats could be considered more procedurally fair in some ways but not in others.

Reactions to Tests: Multiple Dimensions of Fairness

Applicant reactions models (Arvey and Sackett 1993; Gilliland 1993) suggest that reactions to tests are complex, involving perceptions of multiple aspects of tests and the testing process. Gilliland describes ten dimensions of procedural justice and three dimensions of distributive justice that determine applicants’ perceptions of the fairness of the entire selection system. Similarly, Arvey and Sackett note several dimensions that influence applicant perceptions. Research has generally supported these applicant reactions models. For example, Rynes and Connerley (1993) found that college students generally preferred tests with high apparent content validity (e.g., simulations) over tests with more abstract content (e.g., cognitive ability tests). Smither et al. (1993) compared several selection procedures in terms of their perceived job-relatedness among a sample of applicants for managerial jobs. They found that methods such as biodata were perceived as less job-related than more concrete methods such as assessment centers. Similarly, Macan et al. (1994) compared reactions to a cognitive ability test and an assessment center among applicants at a Fortune 500 company. They found that assessment centers were perceived to be more face valid than the cognitive ability test, even if applicants felt they had performed less well on the assessment center. Finally, Chan and Schnitt (1997) compared the reactions of college students to paper-and-pencil and VB versions of a situational judgment test. The group that took the VB test rated it higher in terms of face validity.

However, prior research is limited in three ways. First, most work has not assessed multiple dimensions of fairness simultaneously as suggested by Gilliland’s (1993) model. Moreover, even when considering multiple dimensions of applicant reactions, most research has not been based in comprehensive applicant reactions models (e.g., Smither et al. 1993). Finally, this research has frequently relied heavily on student samples (e.g., Gilliland 1994; Rynes and Connerley 1993).

In the present study we used Gilliland’s (1993) applicant reactions model to explore the multiple dimensions of fairness. Specifically, we explored the reactions to two tests (a VB test and MC test for police recruit) in terms of eight procedural justice rules that applicants would be able to assess at the time of testing. These eight rules fell into two categories: (1) formal test characteristics; and (2) interpersonal treatment by human resources personnel during the selection process. We chose these eight particular dimensions because we felt that they would be salient in most testing situations and could be easily assessed by applicants.

Formal Test Characteristics

According to Gilliland’s (1993) model, one key fairness dimension is job-relatedness, which is the degree to which the selection procedure either appears to measure content that is relevant to the job or appears to be valid. That is, Gilliland’s definition of job-relatedness essentially includes two components, content validity and predictive validity. Empirical support for this differentiation has been found in two previous studies. Smither et al. (1993) found that ten of their items measuring applicant reactions factored into face validity and predictive validity components. Similarly, Bauer et al. (2001) found through an exploratory factor analysis of a scale developed to measure the dimensions of Gilliland’s model that the job-relatedness items actually loaded into two factors, job-relatedness/content and job-relatedness/predictive, leading to an eleven-factor model of fairness.

In the hiring situation for the present study, the MC test was a content valid, paper-and-pencil test, based on a detailed job analysis and input from subject matter experts. Also, its questions generally used a law enforcement context, enhancing its face validity. The VB test (also content valid) presented applicants with videotaped scenarios of law enforcement situations. Applicants responded to open-ended questions, and their responses were audio-taped and scored later by trained raters.

Therefore, both the MC and VB tests used in the present study were content valid, and both used police scenarios to increase face validity. However, because the relevance of the VB test to police work was more observable and concrete, we felt that it should be rated more positively than the written, MC test in terms of both dimensions of job-relatedness (content and predictive). This is also congruent with Smither et al.’s (1993) findings regarding simulations versus abstract test formats in terms of perceived job-relatedness (a combination of face and predictive validity). This leads to the following hypotheses:

Hypothesis 1a: Applicants will have more positive reactions to a content valid VB test.
than to a content valid MC test in terms of job-relatedness/content.

**Hypothesis 1b:** Applicants will have more positive reactions to a content valid VB test than to a content valid MC test in terms of job-relatedness/predictive.

Another key procedural justice rule discussed by Gilliland (1993) is *opportunity to perform*, which involves how much applicants are allowed to demonstrate what they know during selection procedures. Gilliland notes that applicants may perceive interviews as superior to tests in terms of this dimension because interviews allow applicants to communicate directly to the interviewer rather than through test questions. For example, Bies and Shapiro (1988) found that applicant voice in the interview was related to reactions to the interview. Accordingly, Kohn and Dipboye (1998) found that unstructured interviews were perceived more positively than structured interviews and that the dimension of applicant voice was an important consideration. In the present study, because the VB test allowed applicants the freedom to respond orally to open-ended questions about police scenarios (rather than using a more restrictive, structured multiple-choice format), the VB test should be seen as superior to the MC test in terms of opportunity to perform.

**Hypothesis 1c:** Applicants will have more positive reactions to a content valid VB test than to a content valid MC test in terms of opportunity to perform.

However, suggesting that applicant reactions are multi-dimensional, not all research has found that VB tests are perceived as superior to written tests (e.g., Truxillo and Hunthausen 1999). This may be because applicants perceive VB tests as inferior to written tests in terms of certain procedural justice rules. For example, Gilliland (1993) describes *consistency of administration*, or consistency of selection procedures in terms of standardized administration and scoring, as a key procedural justice rule. Support for the importance of the consistency rule to fairness perceptions was found in a study by Ployhart and Ryan (1998), who found that inconsistencies in test administration affected fairness perceptions. In the present study, the VB test required open-ended answers which were recorded to be scored later by subjective (albeit trained) raters. Moreover, the MC test was group-administered, so that applicants could see how others were treated, while the VB test was administered individually, perhaps leaving some room for ambiguity in the minds of applicants about how they were treated. For these reasons, we hypothesized that the VB test would be perceived as less fair than an objectively-scored, MC test in terms of consistency of administration.

**Hypothesis 1d:** Applicants will have more positive reactions to a content valid MC test than to a content valid VB test in terms of consistency of administration.

Gilliland (1993) defines *reconsideration opportunity* as a procedural justice rule involving the degree to which a selection test allows applicants to modify the evaluation process or receive a second chance. Some support for this rule has been found in the context of drug testing, where drug testing was found to be perceived as more justifiable when a second method was used to confirm results (Murphy et al. 1990). In the present study, the VB test required applicants to answer at a set pace for each question and then continue through the rest of the test. In contrast, the MC test allowed applicants to consider and modify their answers. For this reason, we hypothesized that the MC test would be rated as more fair than the VB test in terms of reconsideration opportunity.

**Hypothesis 1e:** Applicants will have more positive reactions to a content valid MC test than to a content valid VB test in terms of reconsideration opportunity.

**Treatment by Human Resources Personnel**

According to Gilliland’s (1993) model, a second group of procedural justice rules falls into the category of treatment by human resources personnel. *Two-way communication* is an interpersonal dimension involving the opportunity for applicants to give input or to ask questions during the selection process. The importance of communication to fairness perceptions has been found in the performance appraisal literature (e.g., Greenberg 1986). Accordingly, Kohn and Dipboye’s (1998) finding that unstructured interviews were perceived more fairly than structured ones would suggest that less structured formats are perceived to be more fair. Because the orally presented, open-ended response format of the VB test examined in this study was more interactive (i.e., requiring oral presentation) than that of the MC test, we also hypothesized that it would be perceived as allowing for more two-way communication.

**Hypothesis 1f:** Applicants will have more positive reactions to a content valid VB test than to a content valid MC test in terms of two-way communication.

Finally, Gilliland (1993) describes two aspects of interpersonal treatment that are relevant to
fairness perceptions. *Propriety of questions* involves whether the human resource staff avoids improper questioning and suppresses personal bias. Accordingly, *interpersonal effectiveness* involves the degree to which applicants are treated with warmth and respect during the selection process. In the present testing situation, we assumed that the quality of interactions with the testing staff was approximately equal for the MC and VB tests. The same testing staff was involved in both tests; moreover, in each case the staff was there to administer the test and answer any questions, but did not directly participate in the test itself. Therefore, although we made no formal hypotheses regarding propriety of questions and interpersonal effectiveness, we compared reactions to the two test formats in terms of these dimensions.

**Research Question 1:** Will applicant reactions to the MC and VB tests differ in terms of propriety of questions?

**Research Question 2:** Will applicant reactions to the MC and VB tests differ in terms of interpersonal effectiveness?

In summary, we believed that test type would interact with fairness dimensions, such that the MC test and VB test would each be rated as more fair on certain of Gilliland’s (1993) dimensions of fairness.

**Relationship Between Procedural Justice Rules and Fairness of the Selection Process**

Gilliland’s (1993) model posits that these individual procedural justice dimensions affect perceptions of the overall fairness of the selection process. Overall fairness in turn affects both personal (e.g., self-efficacy) and organizational (e.g., organizational attractiveness) outcomes. According to the model, the relative weighting of these dimensions in affecting overall fairness is unclear, but it is probably a function of the type of test, the salience of violation of fairness rules, when the fairness dimension is measured during the selection process, and applicants’ past experiences. In any case, the effects of these individual procedural justice rules on the overall fairness of the selection process has remained unexplored in the literature. For example, Bauer et al. (1998) explored the relationship between five dimensions of fairness and applicants’ belief that tests are fair to use in hiring. However, they did not examine effects on the perceived fairness of the organization’s overall selection process.

To address this gap in the literature, the present study used a longitudinal design to explore the extent to which these individual dimensions of fairness are related to the perceived fairness of the overall selection process. Specifically, we examined the relationship between the eight procedural justice rules measured at the time of each test hurdle and the perceived fairness of the organization’s overall hiring process measured after applicants received their test results. This also allowed us to control for applicant outcomes (i.e., whether applicants were allowed to continue in the selection process). We made the following hypothesis:

**Hypothesis 2:** The eight dimensions of the fairness of a MC test and a VB test, measured at the time of testing, will predict the perceived fairness of the organization’s overall hiring process measured after applicants have received their test results and after controlling for applicants’ personal selection outcomes (pass/fail).

**Job-Relatedness and Test-Taking Self-Efficacy**

Gilliland (1993) proposed self-efficacy as a potential outcome in his model, suggesting that procedural justice and distributive justice (outcomes) will interact to affect applicants’ self-efficacy. Specifically, he proposed that distributive justice will affect self-perceptions when procedural justice rules are fulfilled but not when they are violated. Gilliland based this proposition on attribution theory (Weiner 1985), where if the test is perceived as unfair, external attributions for poor performance will be made, but if it is perceived as fair, good performance will be attributed to internal causes. Using students in a simulated hiring situation, Gilliland (1994) found support for this general proposition, such that the job-relatedness of the selection procedure interacted with outcomes. Specifically, when the job-relatedness of the selection procedure was high, job performance self-efficacy increased for participants who were selected, but decreased for those who were rejected using this procedure. When job-relatedness was low, however, the outcome did not affect job performance self-efficacy. Gilliland suggested the importance of self-efficacy in terms of applicant motivation and future search activities. That is, self-efficacy could be important to rejected applicants’ pursuit of future employment.

Bauer et al. (1998) extended this research by exploring the effects of process fairness dimensions and applicant outcomes (i.e., test performance) on test-taking self-efficacy. Specifically, they found that two dimensions of process fairness, information known about the
test and treatment at the test site (interpersonal treatment), interacted with whether applicants passed the test to affect general test-taking self-efficacy. That is, there was a positive relationship between these dimensions of process fairness and test-taking self-efficacy for applicants who passed the test, but a negative relationship for those who failed the test. However, Bauer et al. did not find this relationship for the job-relatedness dimension as would be suggested by Gilliland’s (1994) findings for job performance self-efficacy. In other words, test-taking self-efficacy should have been affected by perceptions of the job-relatedness of the test and test performance, but were not. Furthermore, they did not explore applicants’ test-taking self-efficacies for specific types of tests.

However, the Job-Relatedness × Outcome interaction may not have been found in the Bauer et al. (1998) study due to the measures used. Bauer et al. used a global measure of job-relatedness. As noted, Gilliland (1993) proposed that job-relatedness is best conceptualized as two dimensions, focused either on test content or on test validity, and empirical research has supported these two dimensions of job-relatedness (e.g., Bauer et al. 2001; Smither et al. 1993). Therefore, the relationship between job-relatedness and outcomes such as test performance deserves further study using measures of the two dimensions of job-relatedness. In addition, Bauer et al. (1998) focused on self-efficacy in taking a written, multiple-choice test, a format that would be familiar to most applicants. However, test-taking self-efficacy may be more malleable primarily for test formats that are unfamiliar such as VB tests.

Therefore, we explored whether each type of job-relatedness (content and predictive) would interact with performance on MC or VB tests to affect specific (MC or VB) test-taking self-efficacy. We hypothesized that job-relatedness would be positively related to test-taking self-efficacy for applicants who perform well on the test; the opposite would be found for applicants who perform poorly on the test. Moreover, because applicants should be more familiar, in general, with the MC than the VB test format, we expected that VB test-taking self-efficacy would be more malleable, and that the present study would provide a useful setting for testing these hypotheses.

**Hypothesis 3a:** Job-relatedness/content, measured at the time of the test, will interact with test performance to affect MC and VB self-efficacy after applicants have received their results.

**Hypothesis 3b:** Job-relatedness/predictive, measured at the time of the test, will interact with test performance to affect MC and VB self-efficacy after applicants have received their results.

### Method

**Participants and Procedure**

Participants were recruited from a pool of 379 applicants for an entry-level law enforcement position in a large Southern city. Data for the present study were collected from the same sample as reported in Sanchez, Truxillo, and Bauer (2000), Study 2, except for demographics, variables of interest in the two studies do not overlap. Participation was voluntary and anonymous. Surveys were collected from applicants at five times: before (T1) and after (T2) the first selection procedure, a MC, written test (KR-20 = .91); before (T3) and after (T4) the second procedure, a VB test (interrater reliability = .90), about one month following the MC test; and when applicants received their results indicating whether they were eligible for further consideration (T5), about two months after the VB test, via mail. There was a one-month lag time between the two tests so that the written test could be scored and used as a hurdle if necessary; in this organization, low-scoring applicants are typically dropped from the process and not referred to the subsequent, more expensive selection hurdles (in this case the VB test). However, because scores were such that no applicants were eliminated as a result of the written test, this resulted in a less restricted sample (i.e., those who scored lowest on the MC test were retained in the selection process).

The VB test consisted of scenarios for which applicants orally answered open-ended questions that were audiotaped. Responses were later rated by teams of trained raters and standardized to adjust for differences in rater teams. Although no applicants were eliminated as a result of the written test, applicants had no knowledge of their scores on either test until T5. The test results gave applicants’ scores on each test and their test score band² based on their overall score (i.e., combination of the MC and VB tests). Applicants were also told which bands were likely to be considered for further selection hurdles and possible hiring and which bands were not.

Survey data were matched by the last six digits of the applicants’ social security numbers. Matched data were available for 246 applicants on the first 4 surveys (201 men, 43 women, 2 non-responses; 197 White, 39 African American, 10 other; 65% response rate) and for 100 applicants at all 5 times (78 men, 21 women, 1
Measures

Demographic variables including gender, race, age, and whether the applicant had previously taken a test for this jurisdiction were collected at T1. The perceived fairness of the two tests in terms of eight procedural justice rules suggested by Gilliland’s (1993) model was measured immediately after each test administration (T2 and T4).

The scales in the present study were based on Bauer et al.’s (2001) procedural justice scales, which were developed to tap the process fairness dimensions of Gilliland’s (1993) model. These scales were further refined using exploratory and confirmatory factor analysis. In the present study, we measured eight dimensions of process fairness that applicants would be able to assess at the time of testing. These consisted of job-relatedness/content (2 items; T2 α = .95, T4 α = .96; e.g., ‘It would be clear to anyone that this test is related to the police officer job’); job-relatedness/predictive (2 items; T2 α = .86, T4 α = .86; e.g., ‘A person who scored well on this test will be a good police officer’); opportunity to perform (5 items; T2 α = .91, T4 α = .94; e.g., ‘This test allowed me to show what my job skills are’); consistency of administration (5 items; T2 α = .95, T4 α = .95; e.g., ‘The test was administered to all applicants in the same way’); reconsideration opportunity (5 items; T2 α = .85, T4 α = .94; ‘I am given ample opportunity to have my test results rechecked, if necessary’); two-way communication (5 items; T2 α = .84, T4 α = .93; ‘There was enough communication during the testing process today’); propriety of questions (5 items; T2 α = .91, T4 α = .95; ‘Nothing in the content of the test offended me’); and treatment at the test site (i.e., interpersonal treatment; 5 items; T2 α = .95, T4 α = .96; e.g., ‘I was treated politely during today’s testing process’). Overall fairness of the selection process (based on Smith et al. 1993 and Macan et al. 1994) was measured at T5 (5 items, α = .94; e.g., ‘I think that this testing process is a fair way to select people for the job of police officer’). All were on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). In addition, test-taking self-efficacy (based on Bauer et al. 1998) for the MC test was measured prior to the test (T1; α = .92; 3 items; e.g., ‘I am confident in my ability to do well on written tests’) and when candidates received their results (T5 α = .82). Accordingly, test-taking self-efficacy for the VB test was measured just prior to the test (T3; α = .79; e.g., ‘I am confident in my ability to do well on video tests’) and when candidates received their results (T5 α = .81). Whether the applicant scored high enough on the two tests to be considered for hiring (pass/fail) was determined based on the band into which a candidates’ total score fell.

Results

Applicant Perceptions of Selection Tests

Table 1 presents the means, standard deviations, and intercorrelations for the primary study variables relevant to the MC test. Table 2 presents the means, standard deviations, and intercorrelations for the primary study variables relevant to the VB test. As an overall test of Hypotheses 1a-1e and the research questions, that the MC and VB tests would be rated as differentially fair across the eight justice dimensions, an 8 (Fairness Dimension) × 2 (Test Type) within subjects factorial ANOVA was conducted. Results of the ANOVA indicated a Fairness Dimension × Test Type interaction, F(7, 1694) = 31.00, p < .01, η² = .11. The nature of this interaction was explored through a series of paired samples t-tests, which are presented in Table 3. The t-tests indicated that applicants perceived the VB test as more fair than the MC test in terms of job-relatedness/predictive, opportunity to perform, two-way communication, and treatment at the test site. However, applicants perceived the MC test as more fair than the VB test in terms of consistency of administration and reconsideration opportunity. Finally, the differences in applicants’ perceptions of job-relatedness/content and propriety of questions was nonsignificant. Therefore, Hypotheses 1b, 1c, 1d, 1e, and 1f were supported. With regard to the research questions, no difference was found between the two tests for propriety of questions, but the VB test was rated higher in terms of treatment at the test site.

Response/Non-Response Analysis

All further analyses employed the sample of 100 applicants who completed and returned the final (post-results) surveys. Therefore, we compared differences in the reactions of those who returned earlier surveys and those who did not in terms of the eight fairness dimensions for the MC test (T2), the eight justice dimensions for the VB test (T4), and test-taking self-efficacy prior to testing (T1 and T3). In terms of the eight justice dimensions for the MC test, differences were found between those who responded to the final survey and those who did not for job-relatedness/content, consistency of administration, propriety of questions, and treatment at the test site, (t(208)–2.72), such that those who returned the final survey had responded more favorably to the written test at the time.
Table 1: Descriptive statistics, alpha reliabilities, and intercorrelations for the multiple-choice test

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender</td>
<td>0.18</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Age</td>
<td>27.82</td>
<td>5.95</td>
<td>-0.06</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test Experience</td>
<td>0.50</td>
<td>0.50</td>
<td>-0.03</td>
<td>0.10</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. MC Job-Rel., Content</td>
<td>4.63</td>
<td>0.61</td>
<td>-0.06</td>
<td>0.05</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. MC Job-Rel., Predict</td>
<td>2.64</td>
<td>0.92</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. MC Opp. To Perform</td>
<td>2.81</td>
<td>0.86</td>
<td>0.11</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.10</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. MC Consistency</td>
<td>4.67</td>
<td>0.56</td>
<td>0.01</td>
<td>0.06</td>
<td>0.02</td>
<td>0.54</td>
<td>-0.06</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. MC Reconsider. Opp.</td>
<td>3.61</td>
<td>0.67</td>
<td>0.09</td>
<td>0.17</td>
<td>0.21</td>
<td>0.30</td>
<td>0.23</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. MC Two-Way Comm.</td>
<td>3.91</td>
<td>0.77</td>
<td>-0.03</td>
<td>0.12</td>
<td>0.00</td>
<td>0.31</td>
<td>0.04</td>
<td>0.11</td>
<td>0.44</td>
<td>0.31</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. MC Prop. Of Ques.</td>
<td>4.45</td>
<td>0.64</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.52</td>
<td>0.03</td>
<td>0.11</td>
<td>0.60</td>
<td>0.29</td>
<td>0.53</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. MC Treatment</td>
<td>4.56</td>
<td>0.59</td>
<td>-0.00</td>
<td>0.06</td>
<td>0.07</td>
<td>0.45</td>
<td>0.03</td>
<td>0.07</td>
<td>0.72</td>
<td>0.28</td>
<td>0.48</td>
<td>0.75</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. MC Pre-Test Self-Eff.</td>
<td>3.90</td>
<td>0.71</td>
<td>-0.06</td>
<td>0.02</td>
<td>0.03</td>
<td>0.36</td>
<td>0.02</td>
<td>0.20</td>
<td>0.30</td>
<td>0.22</td>
<td>0.14</td>
<td>0.29</td>
<td>0.27</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. MC Post-Test Self-Eff. (T5)</td>
<td>4.01</td>
<td>0.72</td>
<td>-0.05</td>
<td>-0.04</td>
<td>0.17</td>
<td>0.13</td>
<td>-0.12</td>
<td>0.18</td>
<td>0.04</td>
<td>0.11</td>
<td>0.10</td>
<td>0.25</td>
<td>0.19</td>
<td>0.66</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Overall Fairness (T5)</td>
<td>3.87</td>
<td>0.83</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.12</td>
<td>-0.01</td>
<td>-0.14</td>
<td>0.21</td>
<td>0.09</td>
<td>0.23</td>
<td>0.04</td>
<td>0.11</td>
<td>-0.00</td>
<td>0.18</td>
<td>0.21</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>15. Pass/Fail</td>
<td>0.73</td>
<td>0.44</td>
<td>0.04</td>
<td>0.00</td>
<td>-0.11</td>
<td>0.09</td>
<td>-0.06</td>
<td>-0.04</td>
<td>0.03</td>
<td>-0.14</td>
<td>0.15</td>
<td>0.11</td>
<td>-0.03</td>
<td>0.27</td>
<td>0.28</td>
<td>0.20</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: N = 246 for variables 1-12, where correlations of .12 or greater are significant at p < .05. N = 100 for variables 13-15, where correlations of .21 or greater are significant at p < .05. Pass/Fail = 1 for those who could be considered for hiring; Pass/Fail = 0 for those who could not. For Test Experience, whether applicant had taken a test with this jurisdiction before, 0 = No, 1 = Yes. For Gender, 0 = male, 1 = female.
Table 2: Descriptive statistics, alpha reliabilities, and intercorrelations for the video-based test

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender</td>
<td>0.18</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Age</td>
<td>27.82</td>
<td>5.95</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Test Experience</td>
<td>0.50</td>
<td>0.50</td>
<td>0.03</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. VB Job-Rel., Content</td>
<td>4.66</td>
<td>0.58</td>
<td>-0.11</td>
<td>0.02</td>
<td>-0.02</td>
<td>(0.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. VB Job-Rel., Predict</td>
<td>3.04</td>
<td>0.88</td>
<td>0.16</td>
<td>0.11</td>
<td>0.08</td>
<td>0.12</td>
<td>(0.86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. VB Opp. To Perform</td>
<td>3.13</td>
<td>0.80</td>
<td>0.06</td>
<td>0.14</td>
<td>0.03</td>
<td>0.22</td>
<td>0.54</td>
<td>(0.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. VB Consistency</td>
<td>4.41</td>
<td>0.67</td>
<td>0.06</td>
<td>-0.01</td>
<td>-0.10</td>
<td>0.41</td>
<td>0.19</td>
<td>0.27</td>
<td>(0.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. VB Reconsider. Opp.</td>
<td>3.45</td>
<td>0.75</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.25</td>
<td>0.29</td>
<td>0.25</td>
<td>0.30</td>
<td>(0.94)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. VB Two-Way Comm.</td>
<td>4.35</td>
<td>0.60</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.16</td>
<td>0.40</td>
<td>0.18</td>
<td>0.27</td>
<td>0.61</td>
<td>0.44</td>
<td>(0.93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. VB Prop. Of Ques.</td>
<td>4.52</td>
<td>0.57</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.10</td>
<td>0.48</td>
<td>0.19</td>
<td>0.24</td>
<td>0.63</td>
<td>0.31</td>
<td>0.76</td>
<td>(0.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. VB Treatment</td>
<td>4.66</td>
<td>0.53</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.08</td>
<td>0.45</td>
<td>0.16</td>
<td>0.24</td>
<td>0.60</td>
<td>0.27</td>
<td>0.66</td>
<td>0.76</td>
<td>(0.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. VB Pre-Test Self-Eff.</td>
<td>3.60</td>
<td>0.68</td>
<td>-0.12</td>
<td>0.14</td>
<td>-0.04</td>
<td>0.19</td>
<td>0.30</td>
<td>0.37</td>
<td>0.16</td>
<td>0.26</td>
<td>0.17</td>
<td>0.18</td>
<td>0.20</td>
<td>(0.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. VB Post-Test Self-Eff. (T5)</td>
<td>3.49</td>
<td>0.85</td>
<td>-0.14</td>
<td>0.18</td>
<td>0.18</td>
<td>0.08</td>
<td>0.14</td>
<td>0.30</td>
<td>-0.17</td>
<td>0.06</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.38</td>
<td>(0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Overall Fairness (T5)</td>
<td>3.87</td>
<td>0.83</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.12</td>
<td>0.32</td>
<td>0.14</td>
<td>0.38</td>
<td>0.16</td>
<td>0.23</td>
<td>0.23</td>
<td>0.22</td>
<td>0.14</td>
<td>0.25</td>
<td>0.35</td>
<td>(0.94)</td>
<td></td>
</tr>
<tr>
<td>15. Pass/Fail</td>
<td>0.73</td>
<td>0.44</td>
<td>0.04</td>
<td>0.00</td>
<td>-0.11</td>
<td>0.18</td>
<td>0.00</td>
<td>0.23</td>
<td>0.13</td>
<td>-0.02</td>
<td>0.24</td>
<td>0.12</td>
<td>0.20</td>
<td>0.11</td>
<td>0.06</td>
<td>0.21</td>
<td></td>
</tr>
</tbody>
</table>

Notes: N = 246 for variables 1-12, where correlations of .12 or greater are significant at $p < .05$. N = 100 for variables 13-15, where correlations of .21 or greater are significant at $p < .05$. Pass/Fail = 1 for those who could be considered for hiring; Pass/Fail = 0 for those who could not. For Test Experience, whether applicant had taken a test with this jurisdiction before, 0 = No, 1 = Yes. For Gender, 0 = male, 1 = female.
they took it. In terms of the eight justice dimensions for the VB test, no differences were found between those who responded to the final survey and those who did not, t < 1.47, ns. Finally, no differences were found between those who responded to the final survey and those who did not in terms of either type of test-taking self-efficacy, t < 1.53, ns. Taken together, these findings suggest that the only differences between those who responded to the final survey and those who did not were in terms of four of the fairness measures for the MC test.

Relationship with Overall Fairness

According to Hypothesis 2, the eight fairness dimensions measured at the time of each test would predict the overall perceived fairness of the selection system measured when applicants had received their test results, after controlling for applicants’ personal outcomes (pass/fail). To test Hypothesis 2, two hierarchical regression equations were calculated with overall fairness at T5 as the dependent variable. Gender, age, whether the applicants had taken the test before, and pass/fail were entered on the first step as control variables. For the first equation, the eight justice measures for the MC test were entered on the second step. For the second equation, the eight justice measures for the VB test were entered on the second step.

These regression equations are presented in Tables 4 and 5. The addition of the eight justice rules for the MC test added to the prediction of overall selection system fairness, $R^2 = .18$, $F$
Specifically, beta-weights for job-relatedness/predictive; opportunity to perform; and reconsideration opportunity were statistically significant, although job-relatedness/predictive was in a negative direction. Similarly, the addition of the eight justice rules for the VB test added to the prediction of overall selection system fairness, $R^2 = .19$, $F(8, 84) = 2.73$, $p < .05$. Specifically, beta-weights for job-relatedness/content and opportunity to perform were statistically significant. Therefore, Hypothesis 2 was generally supported.

Finally, we ran an additional regression in which the control variables were entered in Step 1 and all 16 measures of test fairness (8 post-MC and 8 post-VB) were entered on Step 2. These results should be interpreted cautiously because of the substantial loss of degrees of freedom from the large number of predictors relative to cases. However, this analysis revealed a significant increase in $R^2$ on Step 2, $\Delta R^2 = .26$, $F(16, 73) = 2.26$, $p < .05$. Specifically, for the MC test beta weights for opportunity to perform (.24, $p < .10$) and reconsideration opportunity (.24, $p < .10$) were significant; for the VB test, the beta for job-relatedness/content (.28, $p < .05$) was significant.

### Job-Relatedness $\times$ Test Performance Interaction

We used hierarchical regression to test Hypotheses 3a and 3b, that job-relatedness would interact with test performance to affect MC and VB test-taking self-efficacy after candidates had received their test results (T5). We controlled for pre-test levels of test-taking self-efficacy by entering pre-test, specific (MC or VB) test-taking self-efficacy on the first step. Note that this was a conservative test of the hypotheses: Pre-test (T1) and post-results (T5) MC test-taking self-efficacy were very highly correlated ($r = .67$); and pre-test (T3) and post-results (T5) VB test-taking self-efficacy were moderately correlated ($r = .38$).

#### Job-relatedness/content (Hypothesis 3a)

We used two hierarchical regression equations (one for the MC test and one for the VB test) to explore Hypothesis 3a, that job-relatedness/content measured just after testing would interact with test performance to affect specific test-taking self-efficacy measured after candidates received their results (T5). In the first equation, T5 MC test-taking self-efficacy was the dependent variable. Job-relatedness/content (T2) of the MC test, MC test score, and pre-test (T1) MC test-taking self-efficacy (as a control variable) were entered on Step 1. Then the interaction term, which was the product of job-relatedness/content (T2) and MC test score, was entered on Step 2. The interaction was not significant, $\Delta R^2 = .01$, $F(1, 93) = .95$, $ns$. In the second equation, T5 VB test-taking self-efficacy was the dependent variable. Job-relatedness/content (T4) of the VB test, VB test score, and pre-test (T3) VB test-taking self-efficacy (as a control variable) were entered on Step 1. Then the interaction
term, which was the product of job-relatedness/content (T4) and VB test score, was entered on Step 2. The interaction, shown graphically in Figure 1, was significant, $\Delta R^2 = .16$, $F (1, 93) = 24.27$, $p < .01$. Specifically, for those who received a low score ($-1$ SD) on the VB test, there was a negative relationship between perceived job-relatedness/content at T4 and VB test-taking self-efficacy at T5. There was a slightly positive relationship for those who received a high score ($+1$ SD) on the VB test. Therefore, Hypothesis 3a was supported for VB test-taking self-efficacy but not for MC test-taking self-efficacy.

Job-relatedness/predictive (Hypothesis 3b) We used two hierarchical regression equations (one for the MC test and one for the VB test) to explore Hypothesis 3b, that job-relatedness/predictive measured immediately following testing would interact with test performance to affect specific test-taking self-efficacy measured after candidates received their results (T5). In the first equation, T5 MC test-taking self-efficacy was the dependent variable. Job-relatedness/predictive (T2) of the MC test, MC test score, and pre-test (T1) MC test-taking self-efficacy (as a control variable) were entered on Step 1. Then the interaction term, which was the product of job-relatedness/predictive (T2) and MC test score, was entered on Step 2. The interaction was not significant, $\Delta R^2 = .01$, $F (1, 93) = 2.53$, ns. In the second equation, T5 VB test-taking self-efficacy was the dependent variable. Job-relatedness/predictive (T4) of the VB test, VB test score, and pre-test (T3) VB test-taking self-efficacy (as a control variable) were entered on Step 1. Then the interaction term, which was the product of job-relatedness/predictive (T4) and VB test score, was entered on Step 2. The interaction was not significant, $\Delta R^2 = .00$, $F (1, 93) = .06$, ns. Therefore, Hypothesis 3b was not supported for either test.

Discussion

A central purpose of the present study was to illustrate the multidimensionality of the perceived fairness of selection procedures. As hypothesized, applicants perceived the MC test as more fair than the VB test on some dimensions, but perceived the VB test as more fair on other dimensions. This research underscores the importance of using multi-dimensional measures of fairness as suggested by Gilliland’s (1993) model in addition to employing broad measures of test fairness. Because selection methods have multiple facets to them (e.g., format, rules of administration, familiarity to applicants), they can be perceived as fair in terms of certain dimensions but not others. Accordingly, the present study confirms prior findings that the job-relatedness dimension in Gilliland’s (1993) model should be differentiated into ‘content’ and ‘predictive’ components (e.g., Bauer et al. 2001; Smither et al. 1993), as reactions to one may differ from reactions to the other. The fact that job-relatedness/content was equal for both groups is likely due to the very high ratings of both tests on this dimension, perhaps because of the high face validity of both tests. Finally, although a true comparison of MC and VB formats was not possible in this field study due to content and order effects, this research may shed some light on differences in the perceptions of these two test formats. For example, VB tests with open-ended formats may be perceived as more fair than MC tests in terms of opportunity to perform but less fair in terms of consistency of administration.

Another key finding of the present research was that fairness perceptions measured at the time of testing were related to the perceived fairness of the selection system measured months later and after controlling for whether
applicants ‘passed’ or ‘failed’ the selection process. For researchers, this finding confirms the continued use of Gilliland’s (1993) model to explain the overall fairness of a selection process. Moreover, Gilliland’s (1993) model suggests a number of factors such as test type that could affect the relative weight of the fairness dimensions in determining overall fairness. In the present study, the regression analyses suggest that the dimensions of job-relatedness and opportunity to perform of both the MC and VB tests were related to overall fairness. The finding that the job-relatedness dimensions were important to overall fairness confirms Gilliland’s proposition about the importance of job-relatedness to overall fairness. Moreover, test type may have been an important factor in determining the weight of these dimensions to overall fairness: It is not surprising that job-relatedness and opportunity to perform were important in determining overall fairness given that both tests were relatively face valid. Future research should continue to explore what affects the relative weight of dimensions in predicting overall fairness.

For organizations, this research is important because fairness reactions have also been found to be related to such outcomes as organizational attractiveness (e.g., Bauer et al. 1998; Macan et al. 1994) and even proposed to be related to legal action on the part of applicants (e.g., Gilliland 1993; Seymour 1988). Therefore, organizations should also consider applicant perceptions when choosing individual selection procedures, as these perceptions contribute to the perceived fairness of the overall selection process. The present study suggests the importance of job-relatedness of selection processes and making applicants feel they had a chance to demonstrate their skills.

Third, we found partial support for the interaction effects of job-relatedness/content and test performance on specific test-taking self-efficacy. In one test of this hypothesis, job-relatedness/content of the VB test interacted with VB test score to affect post-results VB test-taking self-efficacy. Although these results should be interpreted cautiously, combined with past findings (e.g., Bauer et al. 1998), this interaction suggests that while increasing the perceived fairness of tests may lead to improved organizational outcomes, it may have negative effects on personal outcomes (e.g., test-taking self-efficacy) for those who do not perform well on the test. This effect was not found for MC test-taking self-efficacy, but it may be that MC test-taking self-efficacy is relatively less malleable because MC formats are familiar to most applicants; this is illustrated by the substantial correlation between pre-test and post-test MC test-taking self-efficacy ($r = .67$).

Moreover, the interaction between job-relatedness/predictive and test score was not found for either test, further suggesting that job-relatedness may actually have two components (Bauer et al. 2001; Gilliland 1993; Smither et al. 1993). Accordingly, the interaction between individual dimensions of fairness and test performance provides a fruitful area for future research, both for its effects on self-efficacy and for its effects on other outcomes. For example, it may affect other key factors such as job search behavior and test-taking motivation. Finally, the dilemma of the relationship between fairness and self-efficacy deserves further consideration and research. Specifically, increasing perceived fairness appears to lead to improved outcomes for most applicants, particularly those who do well in the selection process, but for those who do poorly, increased fairness could negatively affect their self-efficacy.

For practitioners, the findings regarding the multidimensional nature of selection test fairness can inform decisions about the development and implementation of selection methods and programs. To state that one particular type of selection method is perceived more positively by applicants than another can over-simplify the reality of the situation. For example, saying that applicants have more positive reactions to simulations than to MC tests might be an over-generalization, because they may see simulations as more fair on certain dimensions (e.g., opportunity to perform) and MC tests as more fair on others (e.g., consistency of administration). Rather, it might be more useful for organizations to consider the particular characteristics of the selection method in assessing how applicants will likely react to it. We recommend, therefore, that organizations striving to improve applicant perceptions focus on the specific dimensions prescribed by applicant reactions models (e.g., Arvey and Sackett 1993; Gilliland 1993) when designing and altering selection systems.

A conclusive comparison of MC and VB tests was not possible in the present research for three reasons. First, the findings are probably most relevant to this particular type of VB test. For example, the VB test in the current study involved an open-ended response format. VB tests in some other research, however, have employed non-open-ended formats (e.g., Chan and Schmitt 1997; Weekley and Jones 1997). It is possible that a MC-type response format would address the relative weaknesses of the VB test found in the present research. Second, due to the field nature of this study, the order of the two tests could not be counterbalanced. However, because the two test administrations were separated by approximately one month, applicants’ ability to make direct comparisons
between the MC and VB tests may have been reduced. Third, the content of the two tests could not be held constant: Administering two tests in reverse order to different applicants or with identical content would be impractical in most actual selection situations. We urge that applicant reactions models (e.g., Gilliland 1993) guide future research on reactions to particular selection test characteristics, particularly for newer types of testing (e.g., telephone or web-based testing).

Some potential limitations of this study should be mentioned. First, as noted, we could not control for content and order differences between the two tests because of the field nature of this study, so that a true comparison of these two test formats was not possible. Also, the significant, negative correlation between job-relatedness/predictive of the MC test and overall fairness (and the corresponding negative beta) should be noted. However, the testing administrator for this organization noted that many applicants who believed in the MC test were particularly unhappy about having to take the VB test. We speculate that some who felt very strongly about the fairness of the MC test were somewhat dissatisfied with the overall process because of the inclusion of the VB test. This finding suggests the complexity of applicant reactions to testing, and this issue deserves further research. In addition, the non-response analysis found that there were some differences between respondents and non-respondents on four of the eight fairness dimensions for the MC test. However, these differences were not found for any of the eight dimensions for the VB test or for either test-taking self-efficacy measure. Future applicant reactions research should continue to assess the effects of non-response. Also, although most of the differences between the fairness dimensions of the two tests were statistically significant, some effect sizes were small. Finally because these were content valid tests appropriate for this particular job, these results cannot be generalized to all MC and VB tests. However, these are formats that are commonly used, and to the extent that other tests are similar in format, administration, and job-relatedness, these results may be generalizable.

By exploring the reactions of actual applicants to two selection methods over time, we have illustrated the multidimensional nature of selection method fairness. Moreover, we have demonstrated how fairness of individual tests measured at the time of testing are related to later perceptions of the selection system and test-taking self-efficacy. We encourage researchers and practitioners to consider theory-based applicant reactions models (e.g., Gilliland 1993) when studying, designing, and altering selection systems.

Notes

1 An earlier version of this study was presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, May, 1999.
2 In test score banding, scores are grouped into ranges called ‘bands’ within which all scores are considered statistically equal; applicants are then chosen from the top band(s) randomly, with preference for protected groups, or using additional selection criteria. See Sackett and Wilk (1994) for a deeper discussion of banding.

References


