The Impact of Justice and Self-Serving Bias Explanations of the Perceived Fairness of Different Types of Selection Tests

N. Schmitt*, F. L. Oswald, B. H. Kim, M. A. Gillespie and L. J. Ramsay
Michigan State University

Reactions to the use of the ACT/SAT, biodata, and situational judgment measures in college admissions decisions were collected from 644 college freshmen. Evaluation of a series of models of fairness perceptions indicated that self-serving bias and organizational justice explanations may both be responsible for these reactions. Examination of respondents' beliefs about their performance compared with other students' performance also elicited responses that may be attributable to concerns about distributive justice. A variety of perceptual processes may explain fairness perceptions, but from a practical perspective, it may be easiest to manipulate examinees' perceptions of the relevance, and indirectly, the perceived fairness of the selection procedures used to make major selection or admissions decisions.

Introduction

In the past decade, personnel selection researchers and practitioners have become interested in the reactions of job applicants to the psychological tests used to make hiring decisions. Perhaps the earliest work in this area was that of Arvey and Sackett (1993); Arvey, Strickland, Drauden, and Martin (1990); and Schmitt and Gilliland (1992). Similar attention to the reactions of examinees was occurring in Europe, where the term *social validity* was used to describe the manner in which people reacted to an organization's employee selection procedures. There are a number of reasons for this heightened interest in reactions to selection procedures (e.g., Schuler & Fruhner, 1993). In general terms, it is important that high stakes selection/admissions decisions be made fairly and that the general public view the use of tests in these situations favorably. These perceptions contribute to the utility and long-term viability of test use. In addition, organizations are apt to be concerned about their reputation in the target labor pool, future sales of company products among rejected employment applicants, potential legal challenges, and possible differences in the measured constructs as a function of test reactions (Schmitt, 2002).

*Address for correspondence: Department of Psychology, Michigan State University, E. Lansing, MI 48824-1117, USA

Background of the Study

Some of the research on reactions to selection procedures (Arvey et al., 1990; Chan, Schmitt, DeShon, Clause, & Delbridge, 1997) has focused on the relationship between test performance and test reactions. Arvey et al., for example, viewed tests as motivators that either diminished or enhanced examinee performance. The implicit theoretical basis of this research stream was that a self-serving bias mechanism (Maas & Volpato, 1989; Miller & Ross, 1975) explained the performance–reactions relationship; in other words, poor-performing examinees reduce the threat to self by evaluating the test as irrelevant to the job and not valid for predicting future successful job performance. Results from these studies show that test performance is positively related to judgments about the test's relevance and test fairness, supposedly because these perceptions serve to maintain an examinee's sense of self-worth or self-esteem. This self-serving bias serves to affect all reactions to tests, including judgments about relevance and fairness.

Gilliland (1993) drew from literature on organizational justice theory (Folger & Greenberg, 1985; Leventhal, 1980; Thibaut & Walker, 1975) to formulate a justice perspective of the nature of various test reactions. This perspective is probably the dominant one in the current study of reactions to selection procedures (Bauer et al., 2001; Colquitt, Conlon, Wesson, Porter, & Ng, 2001).

Gilliland's paper (1993) specifies what and when specific justice principle violations would lead to negative test reactions. Findings from test reactions research generally support the justice perspective. For example, Smither, Millsap, Stoffey, Reilly, and Pearlman (1996) and Smither, Reilly, Millsap, Pearlman, and Stoffey (1993) found that judgments of tests' relevance to jobs (also termed perceived validity in some previous research) for which applicant ability was being evaluated were strongly related to test fairness perceptions. The notion is that one aspect of the selection situation (i.e., college admissions decision in this article) to which examinees are particularly sensitive is the relevance of procedures used to make decisions.

The organizational justice literature distinguishes *procedural* justice, the perceived fairness of the methods used to make organizational decisions, from *distributive* justice, or the perceived fairness of the outcomes of decisions. Colquitt (2001) has also provided evidence for interpersonal justice and informational justice factors that are distinguishable from other procedural justice concerns. Our focus in this paper is on the perceived relevance of the tools used to make decisions, which is only one of the elements of procedural justice. Relevance is the degree to which examinees perceive that the measure is an index of the type of job activities or required knowledge, skills, and abilities that the applicant associates with the job. There are several reasons for this focus. Gilliland (1993) considered perceptions of relevance to be the most important of the procedural justice elements (others include the consistency of the administration of selection procedures, interpersonal treatment issues, and feedback on performance and process), which is reinforced by the findings of Smither and his colleagues (Smither et al., 1993, 1996). Second, variability in relevance perceptions is apt to be larger than other elements in selection situations such as test standardization, administrator performance, and the explanations provided for use of the procedures. These other procedural elements are likely to be constants in such situations.

Chan, Schmitt, Jennings, Clause, and Delbridge (1998) showed that the self-serving bias and organizational justice perspectives are not in conflict with each other, but that both may be contributing to examinee reactions to selection procedures. Using reactions of applicants to reading and video examinations for a State trooper's position, Chan et al. (1998) reported that the perceived job relevance of selection tests affected perceived fairness and that test performance itself affected both perceptions indirectly through perceived performance. These results support both the justice and the self-serving explanations of fairness perceptions.

Purpose of the Study

The present paper replicates and extends the Chan et al. (1998) paper in four important ways. In previous research,

including that of Chan et al. (1998), performance beliefs have been treated as mediators of the relationships between actual performance and relevance and fairness perceptions (see Figure 1). Performance beliefs have been considered a unidimensional construct and have been operationalized as a simple report of examinees' sense that they did or did not perform well on the measure in question. In this study, we assess participants' beliefs regarding performance in two ways, first by asking students how well they thought they performed on the test itself as has been the case in previous research, and second, by asking students how well they thought they performed relative to others taking the test. It was thought that the latter framing of the performance beliefs questions would be more likely to stimulate concerns about relevance and fairness in the high stakes situation in which high-scoring students gain admission to college and others do not. Further, since admissions decisions at most institutions are competitive, examinee beliefs about their relative performance have obvious implications for the outcomes associated with the procedures used. In addition to a heightened likelihood of self-serving bias, the comparative frame of these questions should serve to enhance related concerns about distributive justice.

Chan et al. (1998) reported a lack of evidence for a justice theory prediction that perceived performance (distributive justice) and relevance (procedural justice) interact. This interaction has been found in a wide variety of different situations (Brockner & Wiesenfeld, 1996), though only one (Gilliland, 1994) of the 45 studies Brockner and Wiesenfeld reviewed involved a selection or admission decision, and in that study mixed support for the hypothesized interaction was found. It may be that in these high stakes situations, respondents' reactions to process flaws are not overlooked even when the respondent achieves a desired outcome (the interaction suggests that procedural justice plays a role in determining reactions only when desired outcomes are not forthcoming). Our data and analyses allow the evaluation of this interaction as a determinant of reactions to three different tests. In this study, the comparative frame for performance beliefs should increase the salience of distributive justice concerns so we would expect that evidence for this interaction would be more likely for this measure of performance beliefs.

In this paper, we also examine subgroup differences in perceptions of three different measures. Because of the well-publicized subgroup differences (racial/ethnic and gender) in standardized test scores like the SAT/ACT, it is likely that members of demographic subgroups will be aware of the performance of members of their subgroup and that racial/ethnic minority individuals will perceive the fairness and relevance of the SAT/ACT and other alternative measures quite differently. Gender differences are usually found favoring women on the verbal portions of standardized tests while men do better on the mathematical subtests. While subgroup data on the ACT/SAT are widely

available and familiar, similar data on the biodata and situational judgment inventories (SJI) used in this study are not. Consequently, we hypothesized an interaction between ethnic group and measurement instrument such that differences in perceptions on these measures would be much greater among African and Hispanic Americans than Caucasians. This line of reasoning is similar to arguments presented by van den Bos, Lind, and Wilke (2001). They propose that when a person has information about the outcomes of a comparable other person, this information will be used to assess how to react to his or her outcome. Since the ACT/SAT scores of members of different groups are relatively well known compared to scores on biodata and SJIs, this social comparison explanation of fairness perceptions should predict a relatively larger difference in perceptions of these three measurement instruments for African and Hispanic Americans than Caucasians. Since gender differences are smaller and do not always favor one group over another, we did not propose any hypotheses concerning the reactions of gender groups to these tests.

Finally, the participants in our research knew their ACT/SAT scores at the point when reactions were collected, but they could only guess on their performance on the other two measures. The social comparison explanation of fairness perceptions proposed by van den Bos et al. (2001) also predicts that the role of process fairness (in this paper, perceived relevance) should be less important, and performance perceptions (outcome or distributive fairness) more important in situations in which a person does know what outcome others receive. This latter proposition also led us to predict that judgments about a measure's relevance would play a smaller role in reactions to the ACT/SAT than in reactions to the other two outcomes.

To summarize our objectives, we present data relevant to the following hypotheses:

Hypothesis 1 Comparative performance beliefs will have a larger impact on relevance and fairness perceptions than absolute beliefs of performance levels.

Hypothesis 2 Relevance and performance beliefs will interact in their impact on fairness perceptions, especially when performance beliefs refer to a comparison of one's performance with other test takers.

Hypothesis 3 There will be an interaction between ethnic status and measurement instrument such that differences in reactions to the three measures will be greater for minority group respondents than majority group respondents. Gender differences in reactions to these tests are explored.

Hypothesis 4 Performance beliefs will have greater impact on fairness and less on relevance perceptions regarding the ACT/SAT than the biodata and SJI measures.

The overall model reflecting our hypotheses about the relationships between actual and perceived performance, relevance, and fairness perceptions is depicted in Figure 1 (also presented in Chan et al. (1998)), where actual test performance is thought to influence performance beliefs. In the situation studied here, the examinees would have had evidence of their past performance on the usual standardized tests used in college admissions decisions (i.e., the ACT or SAT) but little information on how they would do on two experimental measures (i.e., biodata and situational judgment measures) designed to assess their motivation
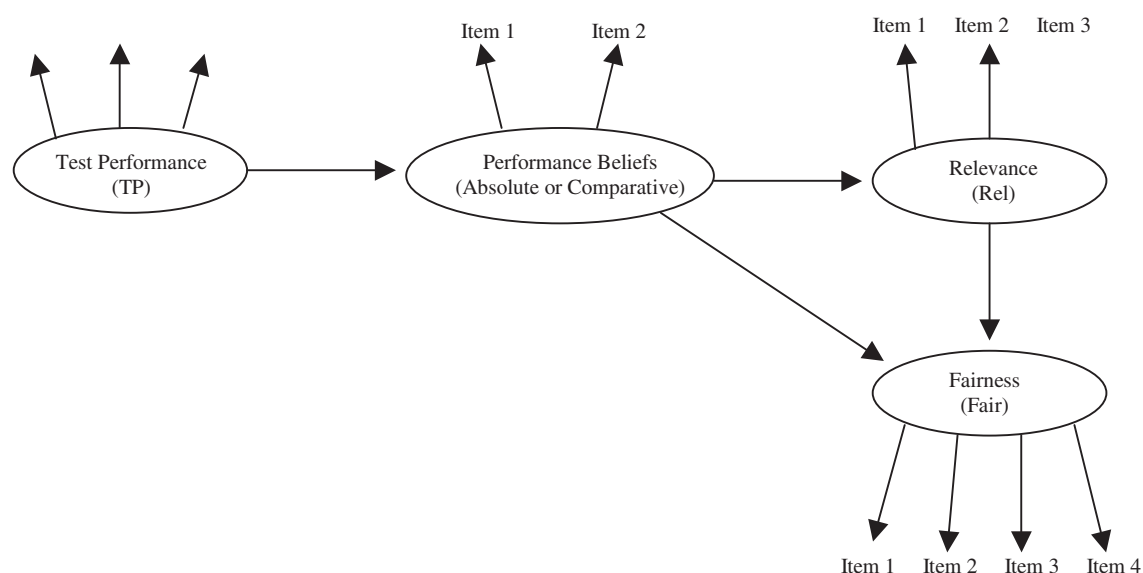

Figure 1. Hypothesized Model of the Relationships of Test Performance, Performance Beliefs, Relevance, and Test Fairness Perceptions

and interest in a set of broad college student performance dimensions. Both actual and perceived performance are included in the model to determine if they have differential direct and indirect impact on test reactions. Individuals will certainly have different perceptions of what constitutes a favorable outcome and it is this evaluation rather than actual performance that will most likely determine their fairness and relevance perceptions. Students will have different standards by which to judge their performance; that is, peers, family members, or what scores are acceptable at various universities. The model indicates that performance beliefs directly influence perceptions of fairness as would be the case if self-serving bias mechanisms were the primary determinant of fairness perceptions, or indirectly through perceptions of relevance, which would be consistent with organizational justice explanations of fairness. However, as indicated above, the proponents of the self-serving bias mechanism hold that this mechanism also affects perceptions of relevance and fairness. In summary, the self-serving bias mechanism is consistent with a prediction that performance beliefs will have both direct and indirect effects on fairness perceptions, while a procedural justice explanation would predict only an indirect effect of performance beliefs on fairness perceptions.

Method

Sample

Six hundred and fifty-four first-year undergraduate students at a large Midwestern university volunteered for this study and received $40 for their participation. Of these, 644 provided usable data after various screens for careless responses. Mean age was 18.5 years (standard deviation = .69). Seventy-two percent were female. The racial/ethnic composition of our sample was 78% Caucasian, 9.5% African American, 2.4% Hispanic American, 5.5% Asian American, and 4.5% other. Freshman students at the university were recruited through their classes, housing units, and through the student newspaper. This sample was very nearly identical to that of the university student body in terms of racial/ethnic identity: 77.3% Caucasian, 9.8% African American, 1.9% Hispanic American, 5.4% Asian American, and the remainder were of other backgrounds. Our sample tended to over-represent females, as 55% of the university's freshmen were women.

Measures

The participants in this research responded to 126 biographical items and 57 situational judgment items measuring motivation, interest, or experience in 12 major college student performance dimensions derived from a review and synthesis of university goal statements, a review of the literature on college student performance, and interviews with university personnel. The dimensions included knowledge acquisition, continuous learning, perseverance, ethics/integrity, social responsibility, leadership, adaptability, interpersonal skills, artistic appreciation, multicultural tolerance and appreciation, physical and psychological health, and career orientation. The biodata and situational judgment items were written to reflect previous experience, interest, or motivation in these 12 areas. Responses to the biographical items provided some support for the existence of the original 12 dimensions, but a general SJI factor accounted for three times the amount of variance accounted for by the second factor and there were no reasonable interpretations of multiple-factor solutions. In the structural equation modeling analyses described in this paper, we formed three biodata and three SJI item parcels. Alphas of the biodata parcels were .83, .79, and .84. Similar values for the SJI parcels were .63, .66, and .63. Items were chosen to represent these parcels randomly. A description of the development of the biodata and SJI measures is available in Gillespie, Kim, Oswald, Manheim, Yoo, and Schmitt (2002) or can be obtained by writing the first author of this paper. Participants also responded to a variety of other interest and personality measures as well as demographic questions.

As mentioned earlier, it should be noted that students were aware of their scores on the ACT/SAT, but did not know their actual performance on the biodata or SJI measures. However, correlations between actual performance on the biodata and SJI indicators and beliefs about their performance were moderate and statistically significant (see the discussion of the results below). Comparison of tests of the hypotheses using measures on which participants knew their actual scores with experimental measures on which participants could only guess how well they did provides evidence for the notion that individuals make self-serving and justice attributions with and without specific feedback on their actual performance.

Also included were measures of the students' performance beliefs, and perceptions of test relevance and fairness. Performance belief items were of two types: one expressing their belief that they had done well on a particular measure, the second that they had done better than others on the measure. Table 1 contains the items addressing all four constructs. Responses to all items were made on five-point Likert-type scales with response options ranging from "Strongly Agree" to "Strongly Disagree." These reaction items were administered after the biodata and SJI items and referred the examinees directly to the measures that they just answered. We also asked them to indicate their reactions to the ACT or SAT using the same items, appropriately reworded to direct their attention to the fact that all students took these standardized measures to gain college admission. Reliabilities of the three fairness measures were .87, .93, and .93; those of the relevance measures, .80, .79, and .80; those of the absolute performance measures, .87, .89, and .93; and those of the comparative performance measures, .84, .88, and .89.

Table 1. Measures of Performance Beliefs, Relevance, and Fairness

Fairness
I feel that this instrument would be a fair way to evaluate college applicants.
Using this instrument to select college students would be fair.
I would be satisfied that I was fairly treated if this instrument were part of the process used to evaluate me for college entry.
Overall, I would be satisfied with the use of this instrument for the selection of college students.

Relevance
I can see a clear connection between this instrument and the activities required of college students.
The actual content of the items in this instrument is related to the types of activities required of college students.
I do not understand what this instrument has to do with the activities required of a college student.

Performance Beliefs: Absolute
I had no problems performing well on this test.
I am confident that I performed well on this test.

Performance Beliefs: Comparative
I am confident that I will be evaluated more highly than others on this test.
I am confident that my performance on the test I just took would be evaluated well above other students taking the test.

All data were collected in group sessions with an average of 15 students (standard deviation = 8). Participants were paid $40 and were given several breaks during the session to avoid respondent fatigue and carelessness. A variety of carelessness items and careful inspection of examinee responses revealed only ten participants whose responses seemed suspicious in some way; all ten were eliminated from our analyses.

Participants were told that the questions we were asking were "measures of judgment and of background experiences and preferences." They were also told that these instruments were "experimental measures that are each linked to what many universities hope will be the outcomes of you attending their university. The major purpose of the project is to test whether the measures of judgment and background are related to your grades and other activities" at the university. They were also informed that we would request that they provide their "reaction to the use of this instrument and other indicators as part of the college admissions process."

Analyses

Latent variable models for biodata, SJI, and ACT/SAT reactions were tested for the two performance constructs resulting in tests of a total of six hypothesized models. In these models, test performance was indexed by three biodata and three SJI indicators. To form these indicators, items were randomly assigned and summed to form one of three composites (see alphas provided above). No measurement model was estimated for the ACT/SAT index since only composite scores were available; that is, a single indicator represented performance on the ACT/SAT. Each of the fairness items was used as an indicator of a fairness latent variable as were the three relevance items and the two sets of performance belief items. These indicators were intercorrelated and these correlations and the variables' standard deviations were used as input to LISREL 8.51 for purposes of assessing the hypotheses depicted in Figure 1 and in our introduction. There were tests of two models, one for each performance construct, for each of the three tests. We also tested models which included direct effects from test performance to relevance perceptions and from test performance to fairness perceptions. Note that these latter models involved a test of effects that are not implied by either the self-serving bias or organizational justice explanations of examinee perceptions of fairness. In evaluating model fit, we used the root mean square error of approximation (RMSEA; Steiger, 1990), the comparative fit index (CFI; Bentler, 1990), and the non-normed fit index (NNFI; Bentler & Bonett, 1980). We employed the usual chi-square difference test to compare models.

To assess the degree to which interactions between relevance and performance beliefs impact fairness perceptions, we used moderated regression analyses. Differences in subgroup perceptions across tests were evaluated using a 4 (ethnic group) by 3 (measurement instrument) analysis of variance with repeated measures on the second factor. Since the four reactions measures were correlated (all less than .45 with the exception of the two performance belief measures assessing reactions to the ACT/SAT which were correlated .76), these analyses of variance were not independent.

Results

Means, standard deviations, and intercorrelations of the test scores and reactions for all three tests examined are

Table 2. Means, Standard Deviations, and Intercorrelations of Biodata and Situational Judgment Instruments

	$\bar{x}$	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Fairness1 (1)	2.54	1.03	1.00													
Fairness2 (2)	2.20	.92	.71	1.00												
Fairness3 (3)	2.51	1.07	.59	.61	1.00											
Fairness4 (4)	2.25	.96	.63	.74	.70	1.00										
Relevance1 (5)	3.25	1.01	.37	.34	.41	.36	1.00									
Relevance2 (6)	3.25	.93	.37	.28	.37	.34	.63	1.00								
Relevance3 (7)	3.39	1.07	.29	.21	.28	.22	.54	.56	1.00							
Performance1 (8)	3.85	.80	.11	.06	.18	.06	.07	.08	.08	1.00						
Performance2 (9)	3.75	.81	.15	.09	.17	.08	.08	.12	.11	.78	1.00					
Comp Perf1 (10)	3.11	.68	.14	.12	.15	.10	.05	.04	.15	.34	.46	1.00				
Comp Perf2 (11)	3.00	.72	.14	.13	.15	.12	.02	−.01	.07	.32	.37	.71	1.00			
Bio (12)	−.21	14.77	.14	.05	.16	.10	.13	.16	.20	.21	.28	.29	.23	1.00		
Bio (13)	−.10	13.16	.11	.03	.16	.08	.16	.19	.24	.15	.21	.26	.22	.83	1.00	
Bio (14)	−.08	15.00	.10	.07	.13	.07	.13	.16	.23	.13	.19	.27	.23	.80	.80	1.00

Note: Comp Perf = Comparative Performance. Listwise deletion yielded N = 598. Correlations above .08 are statistically significant at $p < .05$.

Table 3. Means, Standard Deviations, and Intercorrelations of Biodata and Situational Judgment Instruments

	$\bar{x}$	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Fairness1 (1)	2.24	.96	1.00													
Fairness2 (2)	2.12	.87	.83	1.00												
Fairness3 (3)	2.28	.98	.73	.75	1.00											
Fairness4 (4)	2.09	.86	.73	.81	.78	1.00										
Relevance1 (5)	2.86	1.00	.44	.42	.46	.45	1.00									
Relevance2 (6)	3.06	1.01	.41	.36	.37	.34	.69	1.00								
Relevance3 (7)	3.29	1.03	.25	.17	.26	.20	.51	.53	1.00							
Performance1 (8)	3.57	.77	.02	.00	.06	.07	.16	.13	.09	1.00						
Performance2 (9)	3.52	.76	.03	.03	.07	.05	.17	.13	.05	.05	1.00					
Comp Perf1 (10)	3.03	.62	.09	.14	.15	.15	.11	.07	.06	.13	.11	1.00				
Comp Perf2 (11)	2.97	.63	.04	.08	.09	.11	.12	.01	.06	.12	.09	.07	1.00			
SJI (12)	15.23	7.05	.08	.09	.11	.09	.14	.13	.07	.18	.18	.09	.11	1.00		
SJI (13)	10.08	7.28	.10	.04	.11	.06	.16	.16	.10	.16	.16	.11	.11	.67	1.00	
SJI (14)	12.35	6.98	.11	.03	.11	.04	.18	.18	.11	.16	.18	.10	.12	.66	.65	1.00

Note: Comp Perf = Comparative Performance. Listwise deletion yielded N = 612. Correlations above .08 are statistically significant at $p < .05$.

presented in Tables 2, 3, and 4. As expected, reactions to the use of the ACT/SAT (see Table 4) in college admissions are more highly correlated with actual scores and judgments of examinees' capability to perform well than are similar relationships for the biodata or SJI measures (see Tables 2 and 3). In the case of all three measures, examinees' judgments of their own performance were more highly correlated with actual test performance than were their judgments of test fairness or relevance. Perceptions of relevance were more highly correlated with actual test performance than were judgments of fairness. Interestingly, mean judgments of fairness of the biodata and situational judgment measures were not favorable in that their means were below the midpoint of our five-point scale, but means of the relevance perceptions were above the midpoint of the five-point scale. Fairness perceptions of the ACT/SAT were more favorable than were those for the biodata and SJI. Standardized mean differences between the perceived fairness of these three measures ranged from .5 to 1.2 and all comparisons were statistically significant ($p < .01$). However, the biodata and SJI were seen as more relevant than the ACT/SAT measure. Standardized mean differences on the relevance items were smaller ($d < .4$) and they were not statistically significant ($p > .05$).

Table 4. Means, Standard Deviations, and Intercorrelations of the ACT/SAT and Examinee Reactions

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
Fairness1 (1)	3.06	1.16	1.00											
Fairness2 (2)	3.02	1.11	.82	1.00										
Fairness3 (3)	3.45	1.00	.70	.67	1.00									
Fairness4 (4)	3.12	1.11	.83	.83	.74	1.00								
Relevance1 (5)	3.02	1.08	.50	.49	.46	.51	1.00							
Relevance2 (6)	2.85	1.00	.43	.44	.40	.47	.67	1.00						
Relevance3 (7)	3.30	1.03	.38	.37	.34	.37	.51	.52	1.00					
Performance1 (8)	3.22	1.08	.45	.38	.48	.42	.27	.21	.24	1.00				
Performance2 (9)	3.29	1.06	.47	.41	.52	.46	.27	.24	.27	.87	1.00			
Comp Perf1 (10)	3.01	1.07	.42	.38	.46	.41	.25	.22	.24	.72	.76	1.00		
Comp Perf2 (11)	2.87	1.03	.36	.33	.41	.35	.21	.19	.24	.66	.67	.81	1.00	
ACT/SAT (12)	.55	.68	.41	.33	.44	.38	.19	.14	.19	.62	.65	.63	.54	1.00

Note: ACT and SAT scores were standardized for comparison across the two measures. When multiple tests were available for a single examinee, these scores were standardized and summed. N = 630. Correlations above .08 are statistically significant at $p < .05$.

Table 5. Tests of the Hypothesized Model and Standardized Estimates of Structural Parameters

	ACT/SAT		Biodata		SJI	
	Absol. Perf.	Comp. Perf.	Absol. Perf.	Comp. Perf.	Absol. Perf.	Comp. Perf.
Direct Effects						
TP → PB	.69*	.66*	.26*	.33*	.24*	.13*
PB → Rel	.34*	.32*	.13*	.09	.19*	.11*
Rel → Fair	.53*	.55*	.50*	.50*	.54*	.52*
PB → Fair	.54*	.31*	.07	.13*	− .06	.09*
Indirect Effects						
TP → Fair	.37*	.32*	.04*	.06*	.01	.02*
TP → Rel	.23*	.21*	.03*	.03*	.05*	.02
PB → Fair	.18*	.18*	.04*	.05	.11*	.06*
Fit Statistics						
df	32	32	50	50	50	50
χ^2	102.71	107.03	185.12	174.39	159.28	171.82
RMSEA	.058	.059	.067	.065	.061	.064
CFI	.98	.98	.97	.97	.98	.97
NNFI	.98	.97	.96	.96	.97	.96

Notes: TP = actual test performance, PB = performance beliefs, Rel = relevance, Fair = fairness, Absol. Perf. = absolute performance, and Comp. Perf. = comparative performance.
*$p < .05$.

Model Tests

Table 5 contains results of fitting the theoretical model represented by Figure 1 to the data from all three tests and from the two performance belief measures. The overall fit of all six models to these data was excellent: the non-normed fit indices (NNFI) and comparative fit indices (CFI) were all above .95, the usual value considered to be

evidence of good fit (Hu & Bentler, 1999). In addition, values of the root mean square error of approximation (RMSEA) were all between .06 and .07, also generally considered to be adequate levels of fit (Browne & Cudeck, 1993).

In every model, the parameter relating actual test performance to beliefs about one's performance were large and statistically significant. These parameters were highest

for the ACT/SAT measures as expected, since these students all had access to their own scores. In the case of the biodata and SJI measures, examinees were only inferring their performance from their experience taking the measures without having received feedback on their performance.

More directly relevant to the two theoretical explanations of fairness perceptions are the direct and indirect paths from performance beliefs to fairness. The direct path, which would be consistent with only the self-serving bias explanation, is significant and relatively large for the ACT/SAT examination. This means that students' beliefs or knowledge about how they tend to do on these exams is directly related to test fairness perceptions, as well as being mediated by perceptions of the relevance of the test for college student performance. The mediated relationship, which is consistent with both organizational justice and self-serving bias explanations of the perceptions of fairness, is about one-third as large as the direct effect of performance beliefs on fairness perceptions for ACT/SAT examinations. The fourth hypothesis presented above stated that performance beliefs would have a greater impact on fairness perceptions for the ACT/SAT than for biodata or SJI measures and this is the pattern of results observed in Table 5. The impact of performance beliefs on relevance is smaller than on fairness for the ACT/SAT, but still quite a lot stronger than the similar relationships for the biodata and SJI which is contrary to our expectations.

While the effect of actual test performance on performance beliefs is significant in both models involving biodata, this effect is markedly smaller than the similar effect for the ACT/SAT. Performance beliefs are more likely to lead to perceived relevance than to fairness perceptions, and relevance is highly related to fairness. Indirect or mediated relationships between performance beliefs and fairness are stronger than the direct paths between these two variables when considering absolute beliefs about performance and about equal for comparative beliefs about performance. Taken as a whole, it seems that performance beliefs do lead to judgments about relevance, and these judgments in turn lead to perceptions of fairness.

For the SJI also, the examinees' test performance was directly related to their beliefs about their performance. Relevance judgments were highly related to perceptions of fairness, but the performance beliefs themselves were not highly related to relevance or fairness. Both direct and indirect effects were relatively small, but statistically significant in all but one case, that of the indirect effect from performance beliefs to fairness for the comparative performance construct. In this latter case, there appears to be more support for the self-serving bias perspective in that the direct effect is significant and relatively large. As was speculated in the introduction, framing the performance belief question in this comparative manner did make concern about one's own appraisal more relevant.

There was no evidence to support our first hypothesis that the impact of comparative performance beliefs would have a stronger impact on relevance and fairness perceptions than measures of absolute performance beliefs. Coefficients were comparable or smaller for the comparative beliefs measure than for the absolute measure.

Tests of Alternate Models

While the original models fit the data well, we also evaluated an alternative model for each of the six models in which the direct paths from actual test performance to relevance and fairness were estimated, and compared these models with the original. The original models were nested in the alternative model thus allowing a direct statistical comparison of the two models (i.e., a chi-square difference test with two degrees of freedom). In all six comparisons, the chi-square test was significant, indicating significant direct effects from actual test scores to either judgments about the relevance of the test, or the fairness of the test, or both. However, all changes in the fit statistics were very small (RMSEA <.007, and CFI and NNFI changes of .01 or less). Given that these effects are small or nonexistent, we believe that the perceptual processes involved in deciding on the fairness of these measures is reasonably well captured in the self-serving bias and organizational justice explanations that we explored in this study.

Interactive Effects

In an earlier similar study, Chan et al. (1998) suggested that an alternative interpretation of the perceived performance and fairness perception relationship is that it is a result of distributive justice. Following this line of reasoning, the participants that perform poorly would be upset because they perceive that the admissions outcome would be inequitable. Chan et al. correctly pointed out that the organizational justice literature specifies that procedural justice and distributive justice perceptions interact in their effect on fairness perceptions so that the relationship between procedural rule violation (in this case, a lack of perceived relevance of the procedures used) and fairness is strongest when distributive justice is violated. They found no evidence for this moderator effect. With our second operationalization of performance beliefs in which we ask examinees about their performance compared to other examinees' performance, any violations of distributive justice should be most salient, since students' scores are compared with those of other applicants to college in making admissions decisions. So we predicted that if the interaction between relevance perceptions and performance beliefs that is hypothesized by organizational justice theorists occurs at all, it should be most likely when we examine perceptions about comparative performance as opposed to absolute performance. These notions were tested with a series of moderated regressions in which the sum of the four fairness items was regressed on composites of the relevance perceptions and performance beliefs and,

in a second step of the hierarchical regression, the product of these two predictors. This regression was conducted separately for the two performance variables for each of the three tests. For all three tests, we failed to find evidence for a moderation effect when the performance belief variable involved judgments about one's absolute performance, just as Chan et al. reported. When the performance belief measure involved the examinees' perceptions of their performance as compared with other examinees, however, all three interactive effects were statistically significant ($p < .05$) and consistent with the second hypothesis presented above. That is, the relationships between fairness perceptions and relevance were greatest when performance beliefs, or possible distributive justice concerns were the greatest. Changes in R^2 for all three effects, however, were small ($< .01$).

Subgroup Differences in Test Perceptions

Because information about ethnic subgroup differences on standardized tests is widely available, we hypothesized that test reactions would be partly a function of an interaction between racial/ethnic status and measurement instrument. This hypothesis was tested for each of the four reactions measures using a 4 (ethnic group) by 3 (measurement instrument) analysis of variance with repeated measures on the second factor. Since the four reaction measures were correlated, these analyses of variance are not independent. The means and standard deviations of the four aggregate perceptions by subgroup and measure are provided in Table 6.

For fairness reactions, the measure being considered (ACT/SAT, biodata, and the SJI), race, and the hypothesized interaction were all statistically significant ($p < .05$). As can be seen by examining the subgroup means in Table 6, the perceived fairness of the biodata and SJI measures was roughly similar across subgroups. However, the ACT/SAT was perceived to be significantly fairer by the Caucasian and Asian-American subgroups than it was by the other two racial groups. The measure main effect on fairness perceptions reflected the fact that the overall mean perceived fairness of the ACT/SAT was higher than that of the perceived fairness of the other two measures (2.97 versus 2.44 for biodata and 2.14 for the SJI).

For relevance, only the measure factor was statistically significant ($p < .05$) reflecting the fact that the biodata were perceived to be more relevant than were the ACT/SAT and SJI. The mean relevance judgment for biodata was 3.23 while similar means for the SJI and ACT/SAT were 3.07 and 2.98, respectively.

Perceptions of how well respondents performed in absolute terms were statistically significantly different ($p < .05$) as a function of the measure considered as well as the interaction of measure and race ($p < .05$). Overall,

Table 6. Subgroup Reactions to Biodata, Situational Judgment Inventory, and ACT/SAT Measures

Measure	Group	Fairness Mean	Fairness SD	Relevance Mean	Relevance SD	Performance Belief Mean	Performance Belief SD	Comp. Performance Mean	Comp. Performance Mean
Biodata	Cauc.	2.35	.84	3.22	.84	3.79	.76	3.04	.63
	Afr.Am.	2.42	.90	3.08	.88	3.80	.65	3.03	.61
	His.Am.	2.65	.80	3.29	.92	4.33	.75	3.23	.75
	Asian	2.36	.86	3.25	.82	3.79	.80	3.00	.83
	Male	2.33	.91	2.67	.71	3.76	.74	3.07	.67
	Female	2.39	.85	2.69	.63	3.81	.75	3.04	.65
ACT/SAT	Cauc.	3.25	.97	3.08	.88	3.33	1.03	3.01	1.00
	Afr.Am.	2.65	1.01	2.90	.81	2.97	.97	2.72	.90
	His.Am.	2.68	.99	2.84	.98	2.70	.90	2.66	.96
	Asian	3.29	1.04	3.09	.98	3.00	1.03	2.59	.97
	Male	3.39	.98	2.84	.75	3.40	1.07	3.24	1.02
	Female	3.08	.99	2.57	.71	3.20	1.01	2.84	.96
SJI	Cauc.	2.18	.81	3.09	.85	3.54	.70	2.99	.60
	Afr.Am.	2.19	.81	3.03	.79	3.47	.83	2.90	.58
	His.Am.	2.10	.83	3.22	1.10	3.70	.90	2.90	.28
	Asian	2.10	1.00	2.94	.92	3.60	.83	2.97	.66
	Male	2.12	.86	2.57	.66	3.40	.73	2.95	.69
	Female	2.21	.82	2.60	.66	3.57	.73	2.99	.56

Note: Cauc. = Caucasion, Afr.Am. = African American, His.Am. = Hispanic American, and Asian = Asian American.

respondents felt that they would do significantly better on the biodata (mean = 3.93) and SJI (mean = 3.58) than on the ACT/SAT (mean = 3.00). Considering the interaction of race by measure, an examination of the means in Table 6 reveals that the Caucasian and Asian-American subgroups believed their performance on the ACT/SAT measure was better than did the African-American and Hispanic-American subgroups. In addition, biodata performance reactions on the part of the Hispanic-American group were much superior ($p < .05$) to those of the other three subgroups.

When considering their performance on these three measures relative to other examinees, only the measure effect was statistically significant ($p < .05$). In this case, respondents felt that they would do better on the SJI (mean = 3.70) than on either the biodata (mean = 3.19) or the ACT/SAT (mean = 2.91). The subgroup analyses of test reactions provide strong support for our third hypothesis.

Similar exploratory analyses for gender subgroups revealed a very similar pattern of findings. In the case of all four reactions measures, the repeated measures analysis of variance revealed significant ($p < .05$) measure and gender by measure interactions. The pattern of means in Table 6 for gender subgroups indicates that men react more favorably to the ACT/SAT than do women and that men's reactions to biodata and SJI are either similar or slightly less favorable than are women's reactions to these measures.

Discussion

In this article, we tested a model of test fairness that specified both organizational justice and self-serving bias explanations of fairness perceptions. Results consistently supported the hypothesized relationships specified in the model for all three tests and for both operationalizations of performance beliefs. Applicant perceptions of test relevance are an important correlate of perceptions of test fairness, as would be expected given the organizational justice literature. Tests that applicants in a selection or admissions situation cannot perceive as relevant to the decisions being made represent violations of one of the primary justice rules (Gilliland, 1993). Both relevance and fairness are also directly influenced by performance beliefs, a result consistent with the self-serving bias perspective. This result was particularly strong for the ACT/SAT test for which the students in this study already knew the outcome. These relationships were not as strong, but still statistically significant and relatively large in magnitude, for the biodata and SJI for which the examinees did not have results and could only infer or guess how well they did. The fact that their actual scores on these two measures were relatively strongly correlated with their performance beliefs suggests that these students did have a reasonable sense of how well they performed. The indirect effect of performance beliefs on fairness mediated by relevance

perceptions is consistent primarily with the organizational justice perspective and was significant and modest in magnitude in all but one of the six models.

In terms of the four hypotheses that represent an extension of tests of this model presented earlier by Chan et al. (1998), the results presented did not support our first hypothesis that comparative performance beliefs would have a greater impact on relevance and fairness than would absolute performance beliefs. The framing of these questions seemed to make very little difference in the paths from performance beliefs to these other reactions measures. Positive support was found for the hypothesis that relevance and performance beliefs would interact in their effect on fairness when comparative performance as opposed to absolute performance is considered. These interaction effects were relatively small, but they represent one of the few confirmations that the theoretically hypothesized interaction between procedural and distributive justice actually occurs in a selection situation (Gilliland, 1994; Ryan & Ployhart, 2000). Apparently when the implications of one's score for valued outcomes are salient, the interaction between relevance and fairness perceptions may occur. We also found support for the third hypothesis that members of minority groups would react relatively less favorably to standardized tests as opposed to biodata and SJI than would majority group members. Unexpectedly we found a similar pattern for gender differences; women reacted less favorably than men to the ACT/SAT and both gender groups reacted similarly to biodata and SJI measures. Finally, we did find partial support for the fourth hypothesis in that performance beliefs had a much greater impact on fairness when respondents were considering the ACT/SAT than when they considered the biodata or SJI indices. However, there was little support for the notion that performance beliefs were differentially related to relevance perceptions across the three measures.

Implications for Practice

In an applied situation, it is also appropriate to ask why it would be important to understand the impact of the process fairness and self-serving bias (or distributive justice) explanations of test fairness perceptions. We believe that perceptions of justice are more malleable than are self-serving bias explanations. An individual's satisfactory explanation of the rationale for the procedures being used would presumably have an impact on the perceived fairness of the organization's policies. It may be more difficult, however, to change a self-serving bias. If a person's perceptions of the fairness of a procedure are determined by the outcomes (or expected outcomes) associated with that procedure (Truxillo & Bauer, 1999), then it may be impossible to change those perceptions unless the outcome can be changed. In either event, knowing the psychological mechanisms that influence perceptions provides guidance to those interested in

developing interventions to change those perceptions. Organizational personnel should find it easier to manipulate the relevance of their selection procedures and provide explanations of how the test is relevant to their decisions than it would be to manipulate a self-serving bias. The latter may be an inevitable outcome of the admission or selection decision for which the test is administered. Those who receive a negative evaluation are apt to feel disappointed and attribute some of their misfortune to the selection instruments used to make a decision (Truxillo & Bauer, 1999), but the interaction between distributive justice and relevance suggests that even these people are apt to be influenced positively if the procedures used are seen as relevant to the decision that was made. This finding has potential implications in a wide variety of high stakes testing situations (Sackett, Schmitt, Ellingson, & Kabin, 2001).

Subgroup differences in perceptions of high stakes tests have social consequences in any multicultural society. Our results indicate that reactions to the ACT/SAT test on the part of all groups were relatively favorable; all subgroups perceived the biodata and SJI as less fair than they did the ACT/SAT. However, as predicted, the reactions of Hispanic-American and African-American subgroups were less favorable with respect to the fairness of the ACT/SAT than were Asian-American and Caucasian subgroups. Interestingly, however, the biodata measures were perceived to be the most relevant across groups. There are a number of potential reasons for students' perceptions of the biodata and SJI. They did not know specifically how well or badly they performed; the measures were novel with respect to use in college admissions; and relatedly, it was likely not obvious to the respondents how the items in the SJI items were scored. However, this pattern of results produces an interpretive quandary. If people view the biodata and SJI as more relevant than the ACT/SAT and also view their performance as better on these measures, both self-serving bias and organizational justice explanations suggest that they should rate these measures as fairer than the SAT. A related issue we were unable to address was whether these perceptions might change after the examinees had scores on the SJI and biodata. A future study in which all reactions data are collected prior to the time respondents receive their scores on any of the tests may resolve this confusion. A longitudinal study in which pre- and post-feedback measures are collected would also provide interesting data about the relative role of self-serving and process explanations. Finally, as one reviewer noted, we may also have unintentionally cued our respondents as to the relevance of biodata and SJI measures to college student success with the directions we gave them (see the Methods section of the paper). These instructions were meant to motivate respondents, but it may be best to avoid the possibility of cuing effects in future similar research.

Limitations of the Research

The research described in this paper has a number of limitations. First, it involved respondents for whom the admissions decision had already been made. While all respondents were compensated well for their time, appeared to take the situation seriously, and provided apparently good data, it is true that no decision was being made on the basis of these measures. Although our results are consistent with our hypotheses as depicted in Figure 1, future research should involve respondents in an actual selection or admissions situation. Second, the research includes the evaluation of only one procedural issue. While relevance is arguably the most important, it is certainly not the only issue that may influence judgments regarding the fairness of selection or admissions decisions (Gilliland, 1993). It may also be useful to consider interpersonal and informational justice (Colquitt, 2001) independent of traditional procedural justice issues. How one is treated interpersonally appears to be critically important at various levels of the complete process of organizations' selection and recruitment processes (Rynes, 1993). Third, as noted at various places in the paper, test type is confounded with knowledge of results in this research. Differences in reactions to the different measures may be the result of several factors and future research can be usefully directed to unraveling their relative importance.

References

Arvey, R.D. and Sackett, P.R. (1993) Fairness in selection: Current developments and perspectives. In N. Schmitt and W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 171–202). San Francisco, CA: Jossey-Bass.

Arvey, R.D., Strickland, W., Drauden, G. and Martin, C. (1990) Motivational components of test taking. *Personnel Psychology*, **43**, 695–716.

Bauer, T.N., Truxillo, D.M., Sanchez, R.J., Craig, J.M., Ferrara, P. and Campion, M.A. (2001) Applicant reactions to selection: Development of the Selection Procedural Justice System (SPJS). *Personnel Psychology*, **54**, 387–419.

Bentler, P.M. (1990) Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238–246.

Bentler, P.M. and Bonett, D.G. (1980) Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, **88**, 588–606.

Brockner, J. and Wiesenfeld, B.M. (1996) An integrative framework for explaining reactions to decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin*, **120**, 189–208.

Browne, M.W. and Cudeck, R. (1993) Alternative ways of assessing model fit. In K.A. Bollen and J.S. Long (Eds.), *Testing structural equations models* (pp. 136–162). Newbury Park, CA: Sage.

Chan, D., Schmitt, N., DeShon, R.P., Clause, C. and Delbridge, K. (1997) Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test taking motivation. *Journal of Applied Psychology*, **82**, 300–310.

Chan, D., Schmitt, N., Jennings, D., Clause, C.S. and Delbridge, K. (1998) Applicant perceptions of test fairness: Integrating justice

and self-serving bias perspectives. *International Journal of Selection and Assessment*, **6**, 232–239.

Colquitt, J.A. (2001) On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, **86**, 386–400.

Colquitt, J.A., Conlon, D.E., Wesson, M.J., Porter, C.O.L.H. and Ng, K.Y. (2001) Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, **86**, 425–445.

Folger, R. and Greenberg, J. (1985) Procedural justice: An interpretive analysis of personnel systems. *Research in Personnel and Human Resources Management*, **3**, 141–183.

Gillespie, M.A., Kim, B.H., Oswald, F.L., Manheim, L.J., Yoo, T. and Schmitt, N. (2002). *The development and validation of biographical data and situational judgment inventories in the prediction of college student success*. Presented in symposium titled Beyond g: Expanding thinking on Predictors of College Success at 14th Annual Convention of the American Psychological Society, New Orleans, LA.

Gilliland, S.W. (1993) The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, **18**, 694–734.

Gilliland, S.W. (1994) Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology*, **79**, 691–701.

Hu, L. and Bentler, P.M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1–55.

Leventhal, G.S. (1980) What should be done with equity theory? New approaches to the study of fairness in social relationships. In K.J. Gergen, M.S. Greenberg, and R.H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). New York: Plenum.

Maas, A. and Volpato, C. (1989) Gender differences in self-serving attributions about sexual experiences. *Journal of Applied Psychology*, **19**, 517–542.

Miller, D.T. and Ross, M. (1975) Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, **82**, 313–325.

Ryan, A.M. and Ployhart, R.E. (2000) Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, **26**, 565–610.

Rynes, S.L. (1993) Who's selecting whom? Effects of selection practices on applicant attitudes and behavior. In N. Schmitt and W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 240–274). San Francisco, CA: Jossey-Bass.

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

Schmitt, N. (2002) Do reactions to tests produce changes in the construct measured? *Multivariate Behavioral Research*, **37**, 105–126.

Schmitt, N. and Gilliland, S.W. (1992) Beyond differential prediction: Fairness in selection. In D. Saunders (Ed.), *New approaches to employee management* (pp. 21–46). Greenwich, NJ: JAI Press.

Schuler, H. and Fruhner, R. (1993) Effects of assessment center participation on self esteem and on evaluation of the selection situation. In H. Schuler, J. Farr and M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 109–124). Hillsdale, NJ: Erlbaum.

Smither, J.W., Millsap, R.E., Stoffey, R.W., Reilly, R.R. and Pearlman, K. (1996) An experimental test of the influence of selection procedures on fairness perceptions, attitudes about the organization, and job pursuit intentions. *Journal of Business and Psychology*, **10**, 297–318.

Smither, J.W., Reilly, R.R., Millsap, R.E., Pearlman, K. and Stoffey, R.W. (1993) Applicant reactions to selection procedures. *Personnel Psychology*, **46**, 49–76.

Steiger, J.H. (1990) Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173–180.

Thibaut, J. and Walker, L. (1975) *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Truxillo, D.M. and Bauer, T.N. (1999) Applicant reactions to test score banding in entry-level and promotional contexts. *Journal of Applied Psychology*, **84**, 322–339.

Van den Bos, K., Lind, E.A. and Wilke, H.A.M. (2001) The psychology of procedural and distributive justice viewed from the perspective of fairness heuristic theory. In R. Cropanzano (Ed.), *Justice in the workplace: From theory to practice*. Mahwah, NJ: Lawrence Erlbaum Assoc.