

This is to certify that the thesis entitled

AN INVESTIGATION OF FAKING: ITS ANTECEDENTS AND IMPACTS IN APPLICANT SETTINGS

presented by

ANTHONY S. BOYCE

has been accepted towards fulfillment of the requirements for the

M.A .	degree in	Department of Psychology

Major Professor's Signature

7/18/2005

Date

MSU is an Affirmative Action/Equal Opportunity Institution



THEBIS 2010

PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due. MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
		2/05 p:/CIRC/DateDue.indd-o.1

AN INVESTIGATION OF FAKING: ITS ANTECEDENTS AND IMPACTS IN APPLICANT SETTINGS

By

Anthony S. Boyce

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Department of Psychology

ABSTRACT

AN INVESTIGATION OF FAKING: ITS ANTECEDENTS AND IMPACTS IN APPLICANT SETTINGS

By

Anthony S. Boyce

Researchers have demonstrated that personality-based self-report tests are valid predictors of important organizational criteria including supervisory ratings of job performance, organizational citizenship behaviors, and training performance. However, there remains concern that the validity and utility of such tests may be compromised by intentional distortion, or faking, on the part of applicants. The present study examined both antecedents and consequences of applicant faking using a within-subjects design consisting of the completion of a personality-based selection test at two periods in time. The first administration of the test occurred when participants applied for employment and the second administration occurred under confidential conditions once applicants had been hired. The results indicate that faking is positively related to the extent to which individuals believe that others engage in faking in applicant contexts, but is unrelated to a number of other antecedents investigated. The results also suggest that applicant faking can result in changes in the rank-ordering of individuals. The results do not support a conclusion that faking erodes the criterion-related validity of personality-based tests, but the pattern of results suggests this may be a possibility. The results are discussed in terms of the limitations of the current study and future research directions.

ACKNOWLEDGEMENTS

Thank you to everyone in my life who has ever believed in me, encouraged me, criticized me, or challenged me. Without such, I would surely have found something easier to do than this, but never as satisfying!

Thank you to my advisor, Ann Marie Ryan, for her never ending stream of insightful comments and patience. Thank you to my committee, Drs. Neal Schmitt, Rick DeShon, and John Arnold, for their extremely astute comments and recommendations.

Thank you to my family, specifically my mother and father, Janis and Larry Boyce, and my brother, Jason Boyce, for all their support and unwavering belief in me, even when I had doubts.

Thank you, last but by no means least, to Sarah Conklin for inspiration and allowing me to bore her with details of this thesis more than any person should ever be allowed to bore another.

"There's only one corner of the universe you can be certain of improving, and that's your own self." –Aldous Huxley

iii

L

Ľ

TABLE OF CONTENTS

INTRODUCTION	
The Nature of Faking	
Practical Issues in Faking Research	
Existing Models of Faking	
Past Research	
The Current Model and Study	
Summary	
METHOD	
Sample	
Design	
Measures	
Procedure	
RESULTS	
Construct Validity of the Selection Test	
Selection Test Difference Scores, Scale Reliability, and Effect Sizes	
Descriptive Statistics	
Hypothesis Tests	
Exploratory Analyses	•••••
DISCUSSION	
Does Faking Affect Selection Decisions?	
Are Social Desirability Scales a Valid Operationalization of Faking?	
Does Faking Affect Construct Validity?	
Does Faking Affect Criterion-related Validity?	
Do Individuals' Perceptions & Beliefs Relate to Faking?	
Limitations	
Conclusion	
APPENDIX A	•••••
APPENDIX B	
APPENDIX C	
APPENDIX D	
A DRENIDIV E	

APPENDIX F	
APPENDIX G	
APPENDIX H	
APPENDIX I	
REFERENCES	

LIST OF TABLES

Table 1. Paired Samples t Tests, Effect Sizes, and Reliability Information 5	1
Table 2. Descriptive Statistics and Intercorrelations 52	3
Table 3. Performance Composite Regressed onto Applicant Test Dimensions and Impression Management	7
Table 4. Multiple Groups Confirmatory Factor Analyses 58	8
Table 5. Performance Composite Regressed onto Applicant Test Dimensions Squared . 62	2
Table 6. Rank-order Correlations 64	4
Table 7. D-scores regressed onto Ethics, Beliefs, and Their Interaction 60	6
Table 8. D-scores regressed onto Self-efficacy, Subjective Norms and Their Interaction	8
Table 9. D-scores regressed onto Self-efficacy, Ethics, Beliefs, and Their Interaction 69	9
Table 10. Split-group Correlations and Effect Sizes 72	1
Table 11. Performance Regressed onto Applicant Test Dimension Scores, D-scores, and Their Interaction	4
Table 12. Polynomial Regression onto Applicant and Incumbent Test Dimension Scores And Their Interaction 7'	7
Table 13. Hypothesis Summary	0

LIST OF FIGURES

Figure 1. Model of Faking	. 30
Figure 2. Graph of the Interactive Effects of D-scores and Applicant Test Scores on Performance	. 75
Figure 3. Graph of Polynomial Regression Results	. 78

INTRODUCTION

Personality-based self-report tests are becoming increasingly prevalent in organizational selection processes. These types of measures have proven, across a range of occupations, to be valid predictors of important organizational criteria including supervisory ratings of job performance, organizational citizenship behaviors, and training performance (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Organ & Ryan, 1995). Personality tests have also been shown to exhibit substantially less adverse impact, and to predict similar criteria, than more cognitively loaded measures (Bobko, Roth, & Potosky, 1999; Schmitt, Clause, & Pulakos, 1996). Despite these advantages, there is concern that the validity and practical utility of such tests may be compromised by intentional distortion, or faking, on the part of applicants (e.g., Hogan & Nicholson, 1988; Zerbe & Paulhus, 1987).

The effect of applicant faking on the validity and utility of personality self-report tests has been the subject of a voluminous amount of research in the past decade. Some researchers report evidence that applicants do not fake such measures and that even if faking does occur it does not affect the validity of these instruments (Barrick & Mount, 1996; Hough, 1998; Hough, Eaton, Dunnette, Kamp and McCloy, 1990; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996; Smith & Ellingson, 2002). Conversely, other researchers have found that faking not only occurs and affects criterion-related validity, but affects construct validity and selection decisions as well (e.g., Christiansen, Goffin, Johnston, & Rothstein, 1994; Donovan, Dwight & Hurtz, 2003; Douglas, McDaniel, & Snell, 1996; Ellingson, Sackett, & Hough, 1999; Rosse, Stecher, Miller, & Levin, 1998; Schmit & Ryan, 1993; Topping & O'Gorman, 1997).

There are many potential reasons for these conflicting results. For example, some researchers have relied on the use of social desirability or lie scales to identify fakers in selection settings (e.g., Ellingson, Smith, & Sackett, 2001; Hough, 1998; Rosse et al., 1998). The problem with many of these scales is that it is not clear whether they actually reflect faking or whether they reflect substantive personality traits that have real and meaningful relationships with other traits. Other researchers have used difference scores (d-scores) to operationalize faking behavior (e.g., Dunnette, McCartney, Carlson, & Kirchner, 1962; McFarland & Ryan, 2000). While d-scores are an objective measure of response distortion, research studies that have used these scores were often conducted in laboratory settings where participants are instructed to "fake-good." Such manufactured situations have been useful in showing that individuals can fake, and that faking can impact the validity of personality measures, but fail to show that applicants do fake or that applicant faking does affect validities. Additionally, there is a need for more theoretically-driven research addressing the conditions, both contextual and individual, that lead to faking behavior. Without such research it is difficult to determine why some studies have shown detrimental effects and others have not.

To summarize, despite the interest and efforts of both researchers and practitioners there is still no consensus on whether faking substantially affects the usefulness of personality measures for personnel selection. The study proposed here attempts to inform this debate by addressing some key issues that have been overlooked or under-researched. First, in an effort to address one of the key limitations of past research, the current study will operationalize faking multiple ways (i.e., lie scale and dscores). Second, the current study will attempt to elucidate unexamined, or poorly

examined, proximal antecedents of faking in order to contribute to the theoretical understanding of the conditions that lead to such behavior. Finally, the study described here will utilize a within subjects design to examine the faking of actual applicants in a field setting. Given that the goal of faking research is to generalize findings to real-world applicant settings, it is prudent to examine this phenomenon in such settings.

Before describing the current study in more detail, it is necessary to review the nature of faking and the various ways in which it has been operationalized in the literature. Next applied research investigating the impact of faking on the construct validity, criterion-related validity, and selection decisions that result from personality-based selection tests will be reviewed. In this section, particular attention will be given to the limitations of much of this research for providing definitive conclusions on the effects of faking as well as how the current study will address these limitations. Finally, two models of faking behavior will be reviewed and the model tested in the current study will be presented. It should be noted that faking research has also focused on biodata and situational judgment tests as well, but given the focus of the current research on traditional personality tests, findings from these related literatures will be discussed only when directly applicable to the current research.

The Nature of Faking

Faking, variously termed response distortion, response inflation, and impression management, is a deliberate and conscious attempt to convey false information to create a positive impression on others (Paulhus, 1984; 1986; Zerbe & Paulhus, 1987). Paulhus and other faking researchers (e.g., McFarland & Ryan, 2000; Ellingson et al., 1999)

C

ſ

consider faking as distinct from self-deception, another form of socially desirable responding. Self-deception is an unconscious form of response inflation motivated by a desire to protect one's self from psychological threats and is correlated with healthy psychological traits like self-esteem, high need for achievement, and an internal locus of control (Paulhus, 1986). In accordance with Paulhus' conception, in the remainder of this paper faking, impression management, response distortion, and response inflation will refer to conscious dissembling and not to unconscious forms of socially desirable responding like self-deception.

Some authors have directly considered the motivational processes underlying individuals' desires to present themselves favorably, albeit from a rather macro perspective. Schlenker (1980) posits that at the most general level, an individual's decision to engage in impression management stems from the same motivational sources as most other behavior, that is to maximize expected rewards and minimize expected punishments. Similarly, Leary and Kowalski (1990) suggest that antecedents to engaging in impression management stem largely from three general sources: the goal relevance of impressions, value of desired goals, and discrepancy between the desired and currently perceived image. The goal relevance of impressions refers to the relationship between the desired image and the attainment of social or material outcomes. The value of desired goals encompasses both the importance of goal attainment to the individual as well as the scarcity of the goal. Finally, in Leary and Kowalski's model, discrepancy between the desired and current image refers to both real and imagined divergence in others' impressions of the identity an individual would like to convey. After reviewing research investigating the impact of faking on the use of personality tests in selection, past models

of the behavior that build on the general social psychology theories presented above will be reviewed and the antecedents tested in the current study will be presented.

Practical Issues in Faking Research

While the motivating factors contributing to an individual's decision to fake personality-based selection measures are important, many researchers have been more concerned with the implications of such behavior for the use of these measures in applied settings. Specifically, the major concerns in applied settings surround the issues of response distortion scales used to identify fakers or "correct" trait scores for faking, and examining the effects of faking on construct validity, criterion-related validity, and selection decisions.

Response Distortion Scales. Response distortion scales, variously termed faking scales (Levin & Zickar, 2002) social desirability scales (Crowne & Marlowe, 1960), response validity scales (Hogan & Hogan, 1992), unlikely virtues scales (Hough, 1998), or impression management scales (Paulhus, 1984) have often been examined to determine their effectiveness in identifying fakers or correcting faked trait scores. A number of commercially available tests include such scales (cf. Hough, 1998). These scales typically include items referring to behaviors that are undesirable but extremely common ("I sometimes drive faster than the speed limit") or are very desirable but extremely uncommon (e.g., "I have never dropped litter on the street" Paulhus, 1984). Applicants who score above some pre-set cutoff are considered to have faked their responses.

At first glance, these types of scales seem to function as they should. A metaanalysis by Viswesvaran and Ones (1999) found that social desirability scores were inflated at roughly twice the rate, in difference score terms, as scores on measures of the Big Five when participants were instructed to "fake-good." In another article, the same authors refer to these results as suggesting "... response distortion scales are likely to be useful in flagging individuals who fake," (Ones & Viswesvaran, 1998: p. 249). However, note that only studies utilizing "fake-good" instruction sets were included in the metaanalysis. It is possible that different results would be obtained if studies using "incentivemotivation" instructions or applicant populations were meta-analyzed in a similar fashion. Unfortunately, such a meta-analysis including an effect size for social desirability scales does not yet exist. Additionally, these authors failed to separate response distortion scales measuring self-deception from those measuring impression management, which Paulhus (1984; 1986) argued was a more accurate measure of the conscious dissembling that is faking. Despite the limitations just noted, Viswesvaran and Ones' (1999) meta-analysis does provide some evidence that social desirability scales appear to function in the proposed manner (i.e., groups expected to have higher scores do in fact score higher than groups expected to have lower scores). However, there exist a number of limitations in using these scales to "define" faking in either research or applied contexts.

Perhaps the most severe limitation comes from recent research that demonstrates that responses to these scales reflect more "substance" than "style." Smith and Ellingson (2002) conducted a study in which both applicants and students were given the same personality measures along with three different social desirability scales. Student

administration of the test was conducted under complete anonymity and students were instructed to answer honestly.

To test for differences in method and trait loading across the two groups, Smith and Ellingson utilized multiple-groups confirmatory factor analysis. Assuming that the social desirability scores captured a situation-specific response pattern (i.e., "style") one would expect to see larger method factor loadings (i.e., loadings on the latent social desirability construct) and smaller trait factor loadings (i.e., loadings on the latent personality constructs) in the job applicant group than in the student group. However, the results showed similar method and trait loadings across both groups indicating that the social desirability scales measured substantive trait variance (i.e., "substance") rather than a situation-specific response pattern. Similarly, Hurd and colleagues (2001), using meta-analytic techniques, found that social desirability and personality scale scores shared primarily trait variance in both incumbent and applicant settings.

There also exists evidence that response distortion scales function differently across samples of applicants and non-applicants. Stark and colleagues (2001) conducted a study in which they were able to compare the consequences of faking on construct validity across applicants and non-applicants and across subgroups dichotomized on the basis of impression management scores. Differential test function analyses suggested that the impression management items measured different underlying constructs across groups of applicants and non-applicants. The authors concluded, "This finding casts doubt on the generalizability of research from similar traited faking studies, which compare faking groups created from a single sample of respondents using IM scores" (p. 951).

A third limitation to the use of response distortion scales involves the demonstrated failure of these scales to allow for recovery of honest scores. Ellingson, Sackett, and Hough (1999), utilizing a counterbalanced repeated-measures design, obtained both honest and faked scores from a sample of military personnel on both an unlikely virtues scale and a personality measure. By combining the honest and faked condition scale scores into a single distribution and regressing scores of each personality scale on the unlikely virtues scale, the authors obtained multipliers for each personality scale. Conceptually these multipliers allowed the estimation of scale scores that participants would have obtained if they had exhibited zero intentional distortion in their responses. While this correction effectively removed the standardized mean differences found between the honest and faked scores, the corrected scores, on average, did not correlate with honest scores significantly greater than did faked scores. In addition to the inability of these scales to allow for the recovery of honest scores, there also exists evidence that these scales can result in false-positives where some honest individuals are incorrectly classified as fakers (Zickar & Drasgow, 1996).

Another limitation involves the possibility that responses to these scales reflect a degree of positive mental health. Zerbe and Paulhus (1987) suggest that this is likely to be especially true for those scales that fail to separate self-deception from impression management. As mentioned previously, Paulhus (1986) demonstrated that self-deception is positively related to self-esteem, high need for achievement, and an internal locus of control.

A final limitation is that these scales themselves may be susceptible to faking. As advocated by Whyte (1957) in *The organization man*, test takers opting to respond in a

moderately well-adjusted manner can effectively evade such instruments. Furthermore, Kroger and Turnbull (1975) demonstrated participants were able to evade detection by the MMPI validity scales when coached on how to do so.

The above discussion suggests that social desirability scales are not an adequate operationalization of faking. As such, one would expect these types of scales to be only slightly to moderately correlated with difference scores obtained from a within-subjects administration of a personality test under both motivating and non-motivating circumstances.

<u>Hypothesis 1</u>: Scores on a social desirability scale will have a small to moderate correlation with difference scores.

Construct Validity. Concerns about the influence of faking on construct validity are based on the idea that if measurement equivalence cannot be established across applicant and non-applicant groups then the associations among personality measures and other predictors and measures of job performance may be obscured (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). This is a particularly salient concern given that it is a common practice in many organizations to validate personality-based selection tests on volunteer incumbent samples and then assume that the relationships found will generalize to applicant samples. Research investigating the effect of faking on the construct validity of personality measures indicates that faking can adversely impact construct validity (e.g., Ellingson et al., 1999; Schmit & Ryan, 1993). However, research also indicates that faking does not harm construct validity (e.g., Ellingson et al., 2001; Smith, Hanges, & Dickson, 2001). Some of these differences can be explained by examining the methodologies employed in studying this phenomenon.

Laboratory studies utilizing "honest" and "fake-good" instructional-sets have found the largest and most consistent differences in construct validity (e.g., Douglas et al., 1996; Ellingson et al., 1999). However, many authors have suggested that "fakegood" instructional sets artificially inflate differences beyond what would be expected in applicant settings (e.g., Levin & Zickar, 2002; Ones et al., 1996). Hogan (1991) suggests that impression management requires both the motivation to do so and the ability. This suggests that "fake-good" instructional sets result in larger differences in constructvalidity than examinations conducted in applicant settings because participants are equalized with respect to their motivation under "fake-good" instructions while applicants' motivation to do so is likely to vary. Additionally, Smith and Ellingson (2002) suggest that laboratory studies eliminate some of the "natural deterrents" to response distortion present in applicant settings, such as the fear of being caught. Furthermore, there is evidence that laboratory studies using "fake-good" instructions result in larger standardized mean differences than are commonly observed between applicant and incumbent samples (Birkeland, Manson, Kisamore, Brannick, & Liu, 2003). Thus, "fake-good" instruction sets are useful for showing that faking can impact construct validity, but do not inform whether actual applicant faking does affect construct validity.

Some studies have examined the effects of faking on construct validity by artificially dichotomizing groups on the basis of social desirability scale scores (e.g., Ellingson et al., 2001; Ones & Viswesvaran, 1998). While these studies generally found support for construct validity across groups, as discussed previously the

operationalization of faking as scores on response distortion scales is suspect and is likely to limit the generalizability of conclusions (Smith & Ellingson, 2002; Stark et al., 2001).

Unfortunately, even after removing from consideration those studies with limited validity, the impact of faking on construct validity is still unclear. Studies utilizing applicant samples without dichotomization on the basis of response distortion scale scores will be reviewed next.

Schmit and Ryan (1993) investigated the effects of response inflation on construct validity by comparing a sample of applicants to an employment assistance service to students who took the same test, the NEO-FFI, under non-motivating conditions. Multiple-groups confirmatory factor analysis (MCFA) revealed that the hypothesized five factor solution fit the student sample but not the applicant sample. An exploratory factor analysis revealed a six-factor solution for the applicant sample. Schmit and Ryan suggested that this additional factor represented an "ideal employee" factor as it contained significant factor loadings from composites made up of items referring to being a hard worker, likable, conscientious, courteous, etc. Additionally, the applicant group factor scale intercorrelations were substantially higher than those of the student sample. One possible alternative explanation for the MCFA results is that the samples may have violated the assumption of multivariate normality, discussed below, resulting in the erroneous rejection of a true model for the applicant group.

Similarly to Schmit and Ryan (1993), Weekley, Ployhart, and Harolds (2003) failed to find evidence of measurement invariance across an applicant and incumbent sample. While these researchers did find similar factor forms, evidence of configural invariance, they did not find evidence of similar factor loadings across the two groups, a

minimum condition suggested to be necessary to conclude measurement invariance across groups (cf. Rock, Werts, & Flaugher, 1978). Also similar to Schmit and Ryan (1993), these authors did not present a discussion of the satisfaction or violation of multivariate normality, so it is somewhat unclear whether the results reflect real differences or inflated Type I error.

Smith and Ellingson (2002), discussed previously, conducted MCFA on their sample of applicants for entry-level managerial positions and students. MCFA indicated that the factor form and loadings were not significantly different across the two groups. One reason they suggested for why their findings differed from some prior studies was that past studies relied on estimation procedures that required the assumption of multivariate normality. Violations of this assumption result in inflation of Type I error in proportion to the degree of non-normality in the data set. Given that applicant distributions often violate this assumption (Smith & Ellingson, 2002; Smith et al., 2001), this is a valid critique of prior studies. This study also found similar intercorrelations among the personality dimensions for both groups. One cause for concern in this study was the observation that the student group actually scored higher (i.e., in the more socially desirable direction) on some of the personality dimensions, thus the results could be due to sample-specific irregularities.

Smith, Hanges, and Dickson (2001) conducted another study utilizing procedures robust to violations of multivariate normality. In this study, applicant, incumbent, and student samples were obtained from an archival database maintained by the publishers of the Hogan Personality Inventory. With the exception of the student sample, for which all cases in the database for which there was complete data were used in the analyses, the

samples were randomly selected from the total database, representing many different organizations, to achieve equal sample sizes. Using the student sample to specify the baseline model, MCFA revealed that the hypothesized factor structure actually fit the applicant sample slightly better than the incumbent sample. Furthermore, the model even indicated measurement error invariance across the samples, a result not obtained in any other investigation of the effects of faking on factorial invariance. Intercorrelations of the personality dimensions were also similar across the groups. One drawback of this study is that given the archival nature of the data there was very little information available to adequately describe the samples involved. However, it was noted that incumbents were largely administered the test as part of organization-sponsored self-development and career counseling programs. Whether the administration of the inventory in this context resulted in any motivation on the part of incumbents to appear desirable is unclear.

Two studies examining only applicant populations also speak to this issue of construct validity. One study examined two samples of applicants and found substantially higher intercorrelations among the personality dimensions than a prior study conducted by the same authors on a non-applicant sample using the same scales (Barrick & Mount, 1993; Barrick & Mount, 1996). Collins and Gleaves (1998) also conducted an investigation utilizing only an applicant sample. In this study the personality dimensions were also more highly intercorrelated than expected on the basis of prior research conducted with non-applicant samples. However, confirmatory factor analysis suggested a good fit of the data to the theorized five-factor model.

Two recent studies utilizing item-response theory methodologies for investigating measurement invariance also yield conflicting results. Stark and colleagues (2001),

discussed previously, found the presence of differential item and test functioning (DIF, DTF) across samples of applicants and non-applicants suggesting that faking adversely affects the construct validity of personality scales. Note, however, that the non-applicant sample included respondents who took the inventory for research, counseling, or developmental purposes, so it is difficult to establish that all non-applicants were equally unmotivated to inflate responses.

Robie, Zickar, and Schmit (2001) conducted similar DIF and DTF analyses on a sample of incumbents and applicants within the same organization. These authors found the scales to be more highly intercorrelated on average in the applicant group than in the incumbent group. However, DIF and DTF analyses revealed that these elevated correlations among the scales were not associated with degradation in the psychometric properties of those scales.

To summarize, it appears that applicant and non-applicant responses tend to have similar factor forms (i.e., the same items load on the same factors), but as often as not dissimilar factor loadings, and almost always dissimilar measurement errors across groups. Additionally, applicant groups tend to have higher intercorrelations among scales than non-applicant groups indicating an erosion of discriminant validity among the assessed constructs.

It is interesting to note that there have been no studies that have investigated construct validity using the same sample measured once in an applicant setting and a second time under non-motivating conditions. The combination of a within-subjects design and a field sample has two primary advantages over other types of designs and samples used for examining construct validity issues. First, assuming construct validity

evidence is present for the sample under non-motivating conditions it can be inferred that any deviations found under motivating conditions are due to the context and not due to substantive differences inherent in the samples. Second, a field sample allows assessment of construct validity under real life applicant conditions void of the artificial equalization and potential inflation of respondents' motivation to fake that is present in laboratory settings. Given the equivocal results of past studies investigating the similarity of factor loadings across applicant and non-applicant responses, this issue will be investigated on an exploratory basis in the present study. Unfortunately, the large sample sizes necessary for DIF and DTF analyses prevent the utilization of these methods in the current study's investigation of measurement equivalence across groups.

<u>Hypothesis 2a</u>: Factor forms will not be significantly different across the two measurement periods (i.e., applicant administration and research administration); that is, there will be configural invariance of the personality factors across both administration periods.

<u>Hypothesis 2b</u>: Measurement errors of the personality test will be significantly different across the two measurement periods.

<u>Hypothesis 2c</u>: Average intercorrelations among the scales will be higher when the inventory is administered for application purposes than when it is administered for research purposes.

Criterion-related validity. Many researchers have argued that faking does not substantially affect the criterion-related validity of personality tests and have even gone so far as to call faking the "Red Herring" of personality testing for personnel selection (e.g., Hough, 1998; Ones et al., 1996). However, other researchers have critiqued these

claims and presented evidence to the contrary (e.g., Haaland & Christiansen, 1998; Mueller-Hanson, Heggestad, & Thornton, 2003a). As with the debate over the effects of faking on construct validity, no clear answer emerges.

Faking, often operationalized as scores on a scale purported to measure socially desirable responding, has often been assumed to operate as either a suppressor variable or a moderator. As a suppressor variable, faking is assumed to be positively correlated with predictor scores (i.e., personality measures), but unrelated, or negatively related, to criterion measures (e.g., job performance; e.g., Ones et al., 1996). The resulting effect of such a phenomenon is that in the presence of faking the observed relationship between the personality measure and the criterion is attenuated. Researchers have also suggested that faking may act as a moderator of the relationship between personality scales and various criteria (e.g., Hough et al., 1990). In this situation, the criterion-related validity of a personality measure is expected to change as a function of the degree of faking engaged in by respondents. It has also been hypothesized that faking may act as a mediator or a predictor in its own right (Ones, et al., 1996).

Ones and her colleagues (1996) conducted a meta-analysis to examine the effects of faking on criterion-related validity. These authors investigated the predictor, mediator, and suppressor hypotheses discussed above. Unfortunately, the authors operationalized faking as responses to social desirability scales and thus the results should be interpreted with caution. Social desirability scores did not predict task performance or supervisory ratings of job performance thus precluding the possibility of these scores mediating the relationship between personality and criteria. Additionally, by partialling social desirability scores from personality measures, the authors were able to investigate the

impact of such responding on validity. It was found that social desirability scores did not attenuate the validity of the personality measures. The authors concluded that social desirability did not act as a suppressor variable of this relationship.

Hough et al. (1990) studied faking, again operationalized as responses to a social desirability scale, as a moderator of the relationship between personality and job performance. The study utilized a concurrent-validation sample to examine hypotheses. To test the moderation hypothesis the authors used the mean social desirability score obtained with a separate sample instructed to "fake-good" to dichotomize the validation sample into "overly desirable" and "accurate" responders. While almost a third of the resulting correlations with performance dimensions were significantly different for the "overly desirable" and "accurate" groups, the mean difference between the group correlations was only .03. The authors concluded that socially desirable responding did not moderate the relationship between the personality measures used in this study and performance criteria.

The results of the two studies described above are similar to a number of other studies that have operationalized faking as responses to social desirability scales (e.g., Barrick & Mount, 1996; Christiansen et al., 1994; Ones & Viswesvaran, 1998; Weiner & Gibson, 2000). Similarly to the discussion of construct validity, however, laboratory investigations utilizing "fake-good" instruction sets have yielded conflicting results (e.g., Dunnette et al., 1962; Douglas et al., 1996; Frei, Snell, McDaniel, & Griffith, 1998)

Dunnette and colleagues (1962) study and results are representative of studies conducted using "fake-good" instruction sets. In their study, Dunnette et al. administered the Adjective Checklist to sales employees under both instructions to respond honestly

and instructions to "beat the test." Supervisor performance ratings showed the test to be predictive (correlations for the different dimensions ranged from .22 to .38) for responses obtained in the honest condition. However, when respondents attempted to "beat the test" all test dimensions failed to significantly predict supervisor performance ratings, and indeed some of the dimensions even exhibited correlations in the opposite direction. While "fake-good" instructions have been shown to result in larger differences in observed scores than is observed in applicant conditions, this study, and other studies using similar methodologies (e.g., Christiansen et al., 1994; Frei et al., 1998), clearly shows that faking *can* destroy criterion-related validity.

A more recent study by Mueller-Hanson and her colleagues (2003a) suggests that criterion-related validity is harmed by faking. On the basis of evidence provided by Drasgow and Kang (1984) that correlation coefficients are extremely robust to changes in rank order in only particular ranges of a bivariate distribution, these authors hypothesized that while criterion-related validity may not be significantly different for a motivated group than for a non-motivated group, the validity would be significantly higher for the bottom portion of the motivated group than for the top portion of this group.

To test this hypothesis, Mueller-Hanson and her colleagues conducted a study in which they told one group of participants (i.e., the motivated group) that the personality measure would be used to select people into the next part of the study and that those selected would be eligible for a \$20 cash prize. The criterion measure was performance on a 50-item test that involved simple but time-consuming and tedious exercises. The participants were allowed to quit the performance test whenever they wished with no adverse consequences (i.e., they would still be eligible for the \$20 prize). The

relationship between the personality measure (i.e., an achievement motivation measure) and the criterion was larger for the control group ($\mathbf{r} = .17$, $\mathbf{p} < .05$) than for the motivated group ($\mathbf{r} = .05$, *ns*), but the difference was not significant. When the groups were separated into thirds it was found that the validity of the lower portion of the motivated group distribution ($\mathbf{r} = .45$, $\mathbf{p} < .05$) was significantly greater than the validity of the upper portion of the same group ($\mathbf{r} = .07$, *ns*), while the difference in the control group ($\mathbf{r}_{diff} =$.06) failed to reach statistical significance. The generalizability of the results of this study is bolstered due to the presence of a motivated condition that more closely approximated the conditions of applicant settings. The authors even warned participants in the motivated group of the consequences associated with responding dishonestly (i.e., disqualification from the study and ineligibility for the cash prize). Note that due to the laboratory nature of this experiment and the very narrow criterion used the external validity of the study is somewhat questionable and requires further verification.

Haaland and Christiansen (1998) tested a similar hypothesis, albeit without a control group. In this study, qualified recruits attending a police academy were administered the test prior to being formally offered a space at the academy. The test was not used to select recruits, but rather was forwarded on to local agencies for use in selection of applicants from the pool of graduating recruits. Performance ratings were obtained from police academy officers trained to provide these ratings.

Haaland and Christiansen specifically hypothesized that due to faking, "...one would expect a departure from linearity in construct relationships across different ranges of personality test scores," (p. 3). Indeed, this is exactly what was found. The validity of the test was the same for the entire sample as it was for the lower half of the distribution.

However, for the upper half of the distribution the validity was zero and for the top 15% of scorers the validity was equal to that of the entire sample, although in the opposite direction! When scores were corrected for range restriction these differences became even more pronounced. Assuming fakers were overrepresented in the top 15% of the distribution, it appears that faking *does* impact criterion-related validity, although in a manner more complex than previous researchers realized. Unfortunately, the design of this study prevents the conclusion that the results found were due to faking and not due to some unmeasured factor shared among those at the top of the distribution.

In summary, it appears that faking operationalized as responses to a social desirability scale does not impact criterion-related validity. Furthermore, when examining the entire distribution of scores, it appears the effects of faking on criterion-related validity may be masked due to the insensitivity of correlation coefficients to changes in rank order isolated to only certain areas of the distribution (e.g., the top-end). However, when validity coefficients are examined separately for different ranges of the distribution, there is evidence of deviations from linearity indicating that faking can impact criterion-related validity.

As mentioned previously, operationalizing faking as responses to social desirability scales has many limitations, including the fact that these scales have been shown to reflect true trait variance (Smith & Ellingson, 2002), partialling the responses of these scales from faked responses fails to result in the recovery of true scores (Ellingson, et al., 1999), and the possibility that social desirability scales themselves can be faked (Whyte, 1957). Given these limitations, and others noted previously, it is not surprising that studies examining the influence of faking on criterion-related validity by partialling

social desirability scores from observed scores have failed to show that faking erodes validity.

<u>Hypothesis 3</u>: Partialling social desirability scale scores from applicant personality scores will not result in a significant change in criterion-related validity.

Research discussed previously by Haaland and Christiansen (1998) suggests that the effects of faking on criterion-related validity are masked due to the insensitivity of the correlation coefficient to changes in the rank order at one end of the distribution. The current study uses only selected applicants, who are more likely to be in the upper portions of the distribution; thus, it is expected that significant differences in criterionrelated validity, computed on the entire distribution of scores within the current sample, will be found between the scores obtained in the applicant setting and scores obtained in the research setting. Additionally, given the results found by Haaland and Christiansen (1998) with regards to criterion validity differences across different ranges of the distribution, it is also expected that the current study will find significant differences in validity between the upper and lower halves of the applicant sample score distribution.

<u>Hypothesis 4a</u>: Criterion-related validity will be significantly greater for personality scores obtained in a non-motivating context (i.e., for research purposes) than scores obtained in a motivating context (i.e., for application purposes).

<u>Hypothesis 4b</u>: Criterion-related validity of the upper half of the distribution of applicant sample scores will be less than the criterion-related validity of the bottom half of the distribution.

In relation to the above hypotheses, note that there is evidence that the use of some types of impression management tactics by subordinates positively relates to

supervisory performance ratings, although the effect sizes are generally small (e.g., Gordon, 1996; Wayne & Liden, 1995). Assuming that faking is indicative of an applicant's ability to utilize, and likelihood of actually utilizing, impression management tactics on the job, it could be the case that faking actually increases the validity of personality tests when the criterion is supervisor-rated performance. Two meta-analysis have addressed this possibility (Ones et al., 1996; Viswesvaran, Ones, Hough, 2001). Both of these studies operationalized faking as responses to social desirability scales and found overall correlations between these scales and job performance ratings of less than .05. It is possible that an alternative operationalization of faking may yield different results. However, on the basis of these two meta-analyses and the lack of empirical investigations utilizing an alternative operationalization hypotheses 4a and 4b are proposed.

Selection decisions. Much of the research that has examined the effects of faking on selection decisions and the rank-ordering of applicants suffers from the same limitations discussed above under construct and criterion-related validity. It is interesting to note, however, that some of the same research interpreted as providing evidence that criterion-related validity is not affected by faking also provided evidence that changes in selection decisions and rank-orders of applicants was likely to occur due to faking (e.g., Hough, 1998; Christiansen et al., 1994). For example, Christiansen and his colleagues found that an impression management "correction" suggested that using top-down selection on the basis of raw scores would have resulted in up to 16% of those selected being discrepant hires (i.e., those hired on the basis of raw scores who would not have been hired on the basis of their corrected scores).
Similarly laboratory studies have also found evidence of changes in the rankorder of applicants (e.g., Dunnette et al., 1962; Frei et al., 1998). Mueller-Hanson et al. (2003a), described previously, found that when the motivated and non-motivated groups were combined into the same applicant pool, motivated responders rose to the top of the distribution resulting in an increased likelihood of being selected. In fact, for selection ratios of 60% or less, significantly more motivated group members would have been selected than would be expected on the basis of their representation in the entire applicant pool. The authors concluded that personality tests should only be used to "select-out" the lowest scores instead of being used to "select-in" the top scorers as is done when applicants are selected on a top-down basis. Unfortunately some authors suggest that "select-in" procedures are still commonplace in many organizations (Arthur, Woehr, Graziano, 2001).

Even when these studies are viewed in light of the limitations previously discussed, it is clear that faking can give an advantage to those who choose to engage in such behavior. The current study will only examine those applicants who were hired, indicating that they exceeded minimum cutoffs on the personality test and performed adequately in a pre-hire interview. However, even given this constraint it is likely that some applicants would not have been hired on the basis of their honest responses to the personality measure.

<u>Hypothesis 5a</u>: The rank-ordering of people on the basis of responses obtained in a non-motivating context (i.e., for research purposes) will be substantially different than the rank-ordering of responses obtained in a motivating context (i.e., for application purposes).

<u>Hypothesis 5b</u>: Some of the people hired on the basis of their responses in a motivating context (i.e., applicant setting) would not have been hired on the basis of their responses in a non-motivating context (i.e., research setting).

Existing Models of Faking

The practical concerns surrounding faking are important in their own right, but should be considered within a theoretical framework of what motivates individuals to engage in faking. Snell, Sydell, and Lueke (1999) present an "interactional model" of faking behavior. According to this model, in order for an applicant to successfully fake a noncognitive test the applicant must have both the motivation and the ability to do so. Based on a review of the psychological literature related to dishonest behaviors (e.g., deception, theft, etc.), these authors identify three broad factors hypothesized to influence motivation to fake: demographic, dispositional, and perceptual factors. The authors continue on to provide a laundry list of specific constructs hypothesized to influence each of these factors. While the model presented by these authors is useful as a heuristic map of constructs that may be related to faking, it does not provide the theoretical nesting necessary in order to develop a full explanation of the mediating and moderating variables that influence an individual's motivation to fake. Without an explanation of these mediators and moderators it is very difficult to use this model to investigate ways to detect and deter faking. Additionally, this model fails to account for the ways in which currently known methods of deterring faking operate. For example, warnings not to fake have been shown to deter faking to some extent (Dwight & Donovan, 2002), but this

model does not provide for an explanation of the psychological processes through which this effect occurs.

McFarland and Ryan (2000) present a more thorough model of faking, hypothesizing both mediators and moderators of the process. They suggest that all individual differences related to faking behavior operate through the mediating mechanism of beliefs toward faking, defined as the extent to which an individual holds a belief that faking is an acceptable practice, and the more proximal mediating mechanism of intentions to fake. Situational influences, such as warnings, are hypothesized to moderate the relationship between beliefs toward faking and intentions to fake. While this model is more comprehensive than the Snell et al. (1999) model described above, it is unlikely that beliefs toward faking are the only mediating mechanism through which variables influencing one's motivation to fake operate.

Both Snell et al. and McFarland and Ryan provided models that helped researchers direct their efforts in a more organized and systematic fashion. However, current research findings warrant the investigation of additional antecedents that account for a wider variety of influences and a more complex conception of the ways in which these variables operate to influence one's motivation to fake. For example, research shows that individuals believe that faking on selection tests is not the same as outright lying (Lueke, Snell, Illingworth, & Paidas, 2001). This suggests that a construct capturing an individual's beliefs regarding the similarity of faking and lying should be included in models of faking behavior. Additionally, in order to more adequately explicate the motivational processes that relate to faking, individual-level moderators of such behavior need exploration. For instance, it may be the case that even if an

individual is motivated to fake, his or her self-efficacy for successfully faking may limit the degree that he or she actually engages in response distortion.

The discussion presented below, in order to be comprehensive, will begin by reviewing some of the past research on contextual and individual difference influences on faking behavior that will not be examined in the current study. Next, the model to be tested in the current study will be presented and specific hypotheses will be discussed.

Past Research

Contextual Influences. Contextual influences refers to situational conditions present in the applicant context. These types of influences on faking behavior include warnings not to fake and competition for the job. Dwight and Donovan (2002) conducted a meta-analysis of the literature on warning applicants not to fake and found that these warnings were effective in reducing the degree of response inflation that occurs in applicant contexts. These authors also conducted a follow-up study with college students to determine the types of warnings that were most effective in deterring such behavior. In order to increase generalizability of this study, participants were informed that only the four top scorers would be "selected" to receive a monetary benefit, no other benefits were available (i.e., course credit was not offered for participation). Warnings including both a cautionary note that faking is identifiable and a discussion of the potential consequences of such behavior (e.g., removal from the selection process) exhibited the greatest deterrence effect on actual faking behavior. Thus, it appears that warnings decrease individuals' motivation to engage in faking by decreasing one's belief in being able to fake without being caught and increasing the salience of the consequences that may

follow from such behavior. This is consistent with research suggesting that people are likely to impression manage when either expected benefits of doing so increase or the expected costs of not doing so increase (Schlenker, 1980). Note, however, that warnings only served to reduce, not eliminate, response inflation, indicating that warnings do not represent a panacea.

Perceived competition for a job has also been shown to influence an individual's motivation to engage in faking behavior. Leary and Kowalski (1990) suggest that one's motivation to engage in impression management increases when the desired resource is scarce. Pandey and Rastogi (1979) provide support for this notion in a study in which it was shown that applicants increased their use of the impression management tactic of ingratiation towards an interviewer when perceived competition for the job was high. Lueke and her colleagues (2001) also showed that individuals reported being more likely to fake a personality test when presented with a scenario describing intense competition for a desired job. Competition is likely to influence motivation to fake by increasing an individual's attitude towards the utility of faking, a topic considered more fully in the next section.

Individual Differences. Individual differences in personality represent another set of variables found to influence an individual's motivation to fake. While the list of personality variables that may influence faking is quite long, faking research has largely been concerned with only three: conscientiousness, neuroticism, and Machiavellianism.

Costa and McRae (1985) describe conscientious individuals as responsible and rule-abiding and describe neurotic individuals as being especially concerned with how others view them. Conscientiousness and neuroticism also relate to integrity, indicating

that highly conscientious and less neurotic individuals are generally more honest (Ones, Viswesvaran, & Schmidt, 1993). Additionally, in a laboratory study McFarland and Ryan (2000) found that conscientiousness and neuroticism were related to faking.

Leary and Kowalski (1990) suggest that people high on Machiavellianism are more likely to engage in impression management than those low on this trait. Machiavellianism, defined as a belief that others can be manipulated (Christie & Geis, 1970), relates to self-reported lying and cheating in pursuit of desired ends (Kashy & DePaulo, 1996). Furthermore, Mueller-Hanson and her colleagues (2003b) found that Machiavellianism correlated with difference scores for students taking a personality measure under both "honest" instructions and instructions to respond as if applying for one's "dream job."

The Current Model and Study

The review above highlights the influences of both personality and contextual factors on faking behavior. While these constructs are important, the current study will focus on less researched and more proximal influences on faking behavior. Building on the theories of reasoned action and planned behavior (Ajzen & Fishbein, 1980; Ajzen, 1985) the model presented here attempts to clarify the constructs and psychological processes involved in an applicant's choice of whether or not to engage in faking. The model tested in the current study, presented in Figure 1, retains attitudes, subjective norms, and perceived behavioral control beliefs, or self-efficacy (Ajzen & Madden, 1986), contained in the theory of planned behavior; however, an additional explanatory construct has been added. Specifically, ethical beliefs have been added to the model in

order to capture an influence on motivation that is unique to the context of faking, or deviant behaviors more generally. Additionally, note that the substantive variables in the model are assumed to operate through the mediator of intentions. However, due to the nature of the current study (i.e., a field sample of actual applicants) it is not possible to obtain an uncontaminated measure of intentions to fake prior to the individuals' potential engagement in such behavior.

Attitudes. While not tested in the current study, it is important to highlight the role that attitudes play in the theories of reasoned action and planned behavior as well as the role that attitudes may play in one's motivation to engage in faking. Attitudes towards a given behavior, sometimes called subjective expected utility (Harrison, 1995), are based on both belief strength, the strength of perceived contingencies between performing a behavior and possible consequences of the behavior, and on the valence, or desirability, of those consequences (Ajzen & Fishbein, 1980). Belief strength is the degree to which an individual believes that the consequence will follow from the behavior. Valence is the degree to which one perceives the consequence as desirable or undesirable. Prior research has shown attitudes to predict a variety of volitional behaviors including volunteer attendance (Harrison, 1995), weight loss (Schifter & Ajzen, 1985), and class performance (Ajzen & Madden, 1986).

Although there may be other relevant consequences, the most important with respect to faking are increasing and decreasing one's chance of being hired. Thus an applicant's attitude toward faking is composed of the sum of the products of belief strength and valence for the consequences of: (a) faking and thereby increasing one's chances of selection, and (b) getting caught faking and thereby reducing, or eliminating,





NOTE: Hypotheses are noted by their numbers. Dashed lines indicate variables and relationships not tested in the current study

one's chances of selection. Warnings represent a key construct investigated in prior faking research that is likely to influence faking behavior through its effects on attitudes.

Dwight and Donovan (2002), discussed previously, found that warning applicants that faked scores are detectable and explicitly informing applicants of the consequences (e.g., removal from the selection process) resulted in the largest decrease in response inflation. It is possible that warnings deter faking by increasing both belief strength and negative valence associated with being caught faking. Some authors (e.g., McFarland & Ryan, 2000) argue that warnings moderate the relationship between attitudes towards faking and faking behavior. However, this could be an artifact related to the time of attitude measurement in relation to when warnings are given. For example, if attitudes are measured prior to the time warnings are given it may appear empirically that warnings moderate the relationship between beliefs and faking. However, if attitudes were measured after warnings are given it is likely that the warnings would show a main effect on attitudes by increasing the belief strength that faking will lead to being caught, rather than moderating the relationship between attitudes and faking behavior.

In order to adequately assess attitudes, applicants' attitudes would need to be assessed prior to receiving a selection or rejection decision. The nature of the current study precludes such an investigation.

Subjective Norms. Ajzen and Fishbein (1980) define subjective norms as an individual's perceptions of salient others' beliefs about whether a behavior is, or is not, acceptable. Perceptions of others' beliefs toward a given behavior have been found to predict intentions to perform and subsequent performance of that behavior (e.g., Ajzen & Madden, 1986; Schifter & Ajzen, 1985). These effects have also been found with respect to cheating, lying, and shoplifting (Beck & Ajzen, 1991).

Lueke et al. (2001) assessed individuals' subjective norms in regard to faking and found that respondents who self-reported engaging in response distortion in the past also indicated that they believed that others thought that this behavior was acceptable and appropriate. In another study, Mueller-Hanson and her colleagues (2003b) assessed subjective norms as an indicator of the latent variable perceptions of the situation, which also included belief in the importance of faking and belief in ability to fake as separate indicators. While this measure is not a pure measure of subjective norms, structural equation modeling showed that the subjective norms component was the largest

determinant of perceptions of the situation and, furthermore, that perceptions of the situation were the largest predictor of actual faking.

Although no research has examined the potential of situational influences to change individual's subjective norms, it is possible that the situation can be leveraged to indicate to applicants that faking is not a common or acceptable behavior. For example, test administrators could stress that faking on these tests is the same as lying and that most people do in fact respond honestly to these types of tests. Given the norms against lying in society, such a statement may help to deter faking by causing people to assess salient others' beliefs regarding lying, instead of focusing solely on the perceptions of others' beliefs regarding dissembling on a personality test. While it is unlikely that such a statement would eliminate faking by itself, in conjunction with the traditional warnings of possible detection, it may further decrease applicants' motivation to fake. Therefore, it is important to assess the relationship between subjective norms and faking behavior within an actual applicant sample.

<u>Hypothesis 6a</u>: Perceptions that others believe faking to be an acceptable practice will be related to faking.

While Ajzen and Fishbein (1980) limit their definition to perceptions of others' beliefs, literature suggests that it is appropriate to expand this definition to include perceptions of others' behavior as well. For example, Graham, Monday, O'Brien, and Steffen (1994) found that individuals who believed that a large number of students cheated were more likely to report having cheated themselves. In the literature on faking, similar effects of the perceptions of others' behavior have also been found. For example, Lueke et al. (2001) found that individuals reporting a belief that others distort responses

on selection tests were more likely to report having engaged in faking themselves. Similarly to the discussion of perceptions of others' beliefs, perceptions of others' behavior may be amenable to situational influences aimed at raising applicants' awareness that faking is not a common and acceptable behavior. Thus, it is important to verify prior findings within an actual applicant population.

<u>Hypothesis 6b</u>: Perceptions that others engage in faking will be related to faking.

Ethical beliefs. Neither the theory of reasoned action nor the theory of planned behavior hypothesize an ethical influence on behavior. However, some theorists have argued for the inclusion of a moral-ethical component in models of behavioral decisions (e.g., Etzioni, 1988; Triandis, 1977). Ethics, defined here as an individual's personal beliefs about the inherent goodness or badness of performing a behavior, are separate from instrumental concerns, captured by attitudes, and perceived expectations and behavior of others, captured by subjective norms. Rather, it reflects an internalized pressure to be consistent with one's own value system, void of any social pressures or referents. Ethical concerns have been shown to predict behavior above and beyond attitudes, subjective norms, and self-efficacy (Harrison, 1995).

Leary and Kowalski (1990) suggest that most people have an internalized ethic against lying that prevents them from claiming images blatantly inconsistent with their self-concepts. This notion is supported by a study that utilized the *randomized response technique* to examine faking in which it was found that only 15% of people admitted to giving responses during a selection process that were "completely false or made up" (Donovan et al., 2003). However, 32% reported having "exaggerated my personality characteristics or traits" and 62% reported having "down played what some might

consider my negative attributes." Another study, found that many individuals who admitted to faking in the past, perceived response distortion as "different than lying" (Lueke et al., 2001). Taken together, these studies indicate that although an individual may have an ethical computction against lying, this does not necessarily extend to faking.

Given this disconnect between lying and faking, any assessment of ethical beliefs regarding faking must contain two components. The first must assess the degree to which an individual holds an internalized ethic against lying generally. The second component must assess the degree to which the individual perceives response inflation on a selection test as the same as lying. On the basis of evidence that some people perceive lying as different than faking it is expected that the relationship between ethical beliefs concerning lying and faking will be moderated by beliefs that faking is the same as lying.

<u>Hypothesis 7</u>: The relationship between a self-reported ethic against lying and faking will be moderated by an individual's belief that faking on a selection test is the same as lying, such that in the presence of a belief that faking is not the same as lying the relationship between reporting an ethic against lying and faking will be reduced.

Self-efficacy. Self-efficacy, or perceived behavioral control in the terminology of the theory of planned behavior (Ajzen, 1985; Ajzen & Madden, 1980), has been shown to predict a variety of volitional behaviors including weight loss (Schifter & Ajzen, 1985), task performance (Stajkovic & Luthans, 1998), and volunteer attendance (Harrison, 1995). Self-efficacy refers to the degree to which an individual believes he or she can successfully perform a desired behavior. These beliefs may be based on past experience,

perceived ability, second-hand information obtained from others, and other factors that increase or reduce the perceived difficulty of performing a behavior (Bandura, 1997).

The only study to examine the role of self-efficacy with regard to faking was Lueke et al. (2001). These authors found that self-reported ability to distort one's responses to a personality test was related to self-reported faking behavior in the past. Research has shown that self-efficacy beliefs can be influenced by contextual factors (Bandura, 1997). Given these findings it is possible that self-efficacy regarding faking may be amenable to efforts by test administrators aimed at reducing individuals' beliefs that faking is possible. Thus it is important to establish the influence of such beliefs on applicant faking.

<u>Hypothesis 8</u>: Individuals with high self-efficacy for enhancing their responses to a selection test in a desirable way will be more likely to engage in faking.

Possessing the knowledge of the nature of the construct being assessed may also influence an individual's motivation to fake a noncognitive selection test. Reynolds, Sinar, and Haaland (2003) showed that a pre-testing orientation program describing the nature of the personality constructs measured in a selection test can influence test scores. These researchers compared test scores of applicants receiving a construct-focused orientation program to scores of applicants receiving either no orientation or a general orientation on the format of the test. The group of applicants participating in the construct-focused orientation scored significantly higher on the test than both the group receiving no orientation and the general orientation group.

Frei Snell, McDaniel, and Griffith (1998) measured participants' knowledge of the constructs associated with successful performance in customer service jobs. In an

"applicant" condition in which participants were told to respond as if they were applying for a customer service job, knowledge of the construct significantly related to response inflation as measured by within-subject difference scores. Additionally, prior research has shown that item transparency relates to faking (e.g., Alliger, Lilienfeld, & Mitchell, 1996) suggesting that if individuals are aware of the desirability of the construct being measured, they may be more motivated to fake.

While it is possible that the relationship between knowledge of constructs and faking is fully mediated by self-efficacy, it is more likely that this variable is only partially mediated by self-efficacy. Some minimal knowledge of whether the construct is desirable or not is necessary in order for an individual to enhance his or her responses at all; thus, it is expected that knowledge of constructs will also exert a direct influence on faking.

<u>Hypothesis 9a</u>: Knowledge of the constructs being assessed will be related to faking.

<u>Hypothesis 9b</u>: The relationship between knowledge of the constructs being measured and faking will be partially mediated by self-efficacy.

Bandura (1997) states, "Beliefs of personal efficacy constitute the key factor of human agency. If people believe they have no power to produce results, they will not attempt to make things happen." (p. 3). This suggests that in the absence of some minimum level of self-efficacy for a given course of action, an individual will not even attempt the action. While this is perhaps overstated, it is likely that a lack of self-efficacy greatly diminishes an individual's motivation to pursue an action. In the present context, it is likely that, regardless of other motivating factors, an individual who believes that he

or she is not capable of effectively faking his or her responses to a noncognitive selection test will not attempt to do so.

<u>Hypothesis10a</u>: The relationship between subjective norms and faking will be moderated by self-efficacy, such that in the presence of low faking-related self-efficacy the relationship between subjective norms and faking will be reduced.

For the variables assessing ethical beliefs, the hypothesis below represents a three-way interaction between ethical beliefs regarding lying, beliefs that faking is the same as lying, and self-efficacy.

<u>Hypothesis10b</u>: The relationship between ethical beliefs and faking will be moderated by self-efficacy, such that in the presence of low faking-related self-efficacy the relationship between ethical beliefs and faking will be reduced.

Summary

As noted, some of the above relationships have been investigated in prior studies, however many of these studies have used somewhat weak methodologies that limit generalizability, such as using faking scales to operationalize faking or using a "fakegood" instruction set. While these types of study were useful in the initial stages of faking research, it is time to use more complex and generalizable methodologies utilizing applicant samples with more concrete measures of response distortion.

The greatest contribution of the present study is the utilization of a within-subjects design and a field sample to investigate applicant faking. Researchers suggest that this is the type of design and sample that is required to adequately address the debates in the current literature (e.g., Stark et al., 2001; Mueller-Hanson et al., 2003a; Weekley et al.,

2003). The current study also contributes to the knowledge base of faking by investigating antecedents to faking behavior that have been suggested, but not definitively shown, to influence faking.

METHOD

Sample

The entire sample consisted of 169 part- and full-time employees of a large Midwestern theme park. All employees applied, and were selected, for entry-level positions within the organization between January and July, 2004. Of the entire sample, 9 people had substantial amounts of missing test data and were thus excluded from all analyses. An additional 9 people did not exceed the normative cutoffs, and thus should not have been hired based on their applicant test scores and were not included in the analyses. Due to organizational delays in inserting the social desirability scale into the applicant test, data for this scale was obtained from only 29 participants as applicants, 4 of whom should not have been hired on the basis of their applicant test scores, resulting in an analyzable samples size of 25. Due to a researcher error, the knowledge measure was administered to only 49 participants, 4 of whom should not have been hired on the basis of their applicant test scores, resulting in analyzable sample size of 45. Due to small amounts of other missing data the sample size for all analyses, excluding the social desirability and knowledge measures, is between 147 and 151.

The sample was predominately female (62%). Age data was available for 87% of the sample and indicated that individuals included in the sample ranged between 18 and 72 years of age with a mean of 40. Job title information was available for 86% of the sample and indicated that the sample included food service workers (9%), presenters and tour guides (47%), visitor services and retail employees (21%), security personnel (4%), and custodial workers (3%). Race data was not available for the current sample, but analysis of a large sample of prior applicant data for this organization indicates that applicants are predominately white (53%) or African-American (41%), with smaller

representations of Hispanics (3%), Asian or Pacific Islanders (<1%) and Native Americans (<1%).

Approximately 500 employees were eligible to participate in this study. Applicant test scores were available for 221 employees who were eligible to participate, but did not. Independent-samples t-tests were conducted on applicant test scores to examine any differences betweeen those individuals volunteering for this study and those individuals who did not volunteer. The results confirm that there were no significant mean differences between the current sample and the available sample of eligible employees (for all tests: t (370) < 1.974, n.s).

Design

The experiment utilizes a within-subjects design consisting of the completion of a personality-based selection test at two periods in time. The first administration of the test occured when the participants applied for employment to the organization. The second administration occured 3 - 6 months later between August and November, 2004. Supervisory performance ratings were collected at the end of November, 2004.

Measures

Selection test. The proprietary selection test was developed and validated specifically for this organization. A thorough job analysis was conducted in order to elucidate the personality dimensions important for performance at this organization. Scales were constructed to assess these dimensions, and a concurrent validation study was utilized to establish the validity of the instrument. The final instrument contains five

dimensions: Adaptability, Confidence and Friendliness, Productivity and Quality Focus, Ease of Supervision, and Reasoning and Problem Solving.

Version 1 of the selection test contains 107 self-report items assessing personality constructs, 16 multiple-choice items assessing reasoning ability, and 13 self-report items assessing theft and substance abuse. With the exception of 7 items, all items assessing personality constructs are answered on a 5-point likert-type scale ranging from strongly disagree to strongly agree. Approximately 82% of the full analyzable sample (130 people) took version 1 of the selection test as applicants. Version 2 of the selection test is completely nested (i.e., all of the items in Version 2 were included in Version 1) within version 1 and contains 72 self-report items assessing personality constructs, 16 items assessing reasoning ability, and 13 self-report items assessing theft and substance abuse. Version 2 also contains an additional 17 self-report items assessing response distortion, described subsequently, that are not included in Version 1 of the test and are not scored or used for selection purposes. Approximately 18% of the analyzable sample (29 people) took version 2 of the selection test as applicants. Only the items contained in both versions of the tests will be examined and used to test hypotheses.

In addition to differing in the total number of items, the two versions differed slightly in the percentage of people meeting the minimum criteria for interview eligibility (i.e., Version 1: 72.3%; Version 2: 68.2%), and the average concurrent validity of the dimensions (Version 1: r = .23; Version 2: r = .29). The test is scored, for selection purposes, on an empirically-derived rationally constructed 0 to 3 scale with the 2 least desirable options receiving a score of "0", neutral responses receiving a score of "1", and desirable and highly desirable responses receiving scores of "2" or "3" depending on

whether the options exhibited substantial validity beyond the other options for a given item.

The second administration of the test included only the 72 self-report items from Version 2 and the 17 response distortion items. The theft and substance abuse items were excluded due to both the very sensitive nature of the topics covered by the items and the limited amount of time for the research administration of the test. Additionally, in the interest of time, the reasoning ability multiple-choice items were excluded as these items assess cognitive ability and are not related to the primary hypotheses. Due to the proprietary nature of the selection test, it will not be reproduced here and only a few example items will be provided.

The Adaptability dimension was designed to assess the extent to which individuals flexibly adapt to changes in demands and procedures in the workplace and maintain composure in stressful. Examples of items included in this dimension are, "I enjoy it when I get to do new and different things at work," and, "I'm at my best when I'm challenged and things are difficult."

The Confidence and Friendliness dimension was designed to assess the degree to which an individual enjoys being with others, confidently approaches one-on-one and group interaction situations and is comfortable interacting with both customers and coworkers. Examples of items included in this dimension are, "I often feel uncomfortable around others," and, "I am skilled in handling social situations."

The *Productivity and Quality Focus* dimension was developed to measure the extent to which individuals are detail focused, reliable, responsible, and concerned with the quality of their work. Examples of items included in this dimension are, "I am very

exact in what I do," and, "I almost always do more than is required in work or school activities."

The personality-based items in the *Reasoning and Problem Solving* dimension was designed to assess the extent to which individuals are intellectually curious, creative, and seek out opportunities to learn. Examples of items included in this dimension are, "I avoid reading difficult material," and, "I do not have a good imagination." The test score for this dimension is a composite of the personality and cognitive ability items. In order to replicate how this dimension is used in practice, the cognitive ability items from the applicant administration were used to form the scale score for this dimension in both the applicant and incumbent settings. That is, the cognitive ability items were only administered in the applicant setting and were not administered in the incumbent setting.

The *Ease of Supervision* dimension was developed to measure the degree to which individuals trust supervisors, are willing to take direction, and are generally eventempered. Examples of items included in this dimension are, "A lot of supervisors just enjoy controlling people," and, "I get irritated easily."

Performance Appraisal. The performance appraisal form was developed on the basis of a job analysis and discussions with supervisors (Appendix A). Supervisors completed similar forms for employees who took part in the initial validation of the selection test. Supervisors were informed that this performance appraisal was for research purposes only and would in no way affect employees. The eight performance dimension ratings were averaged to form a composite performance rating that is used for all analyses.

Experimental measures. Efforts were made to use established scales to measure the constructs described below. However, many of the scales used in prior research contained very few items, sometimes as few as a single item (e.g., "beliefs about faking" from Lueke et al., 2001; 2002). Thus, it was necessary to supplement existing scales with additional items in most of the below measures. All scales were assessed using a 5-point likert-type scale ranging from strongly disagree to strongly agree, unless otherwise noted.

The order of the measures presented to participants was the same order in which the measures are discussed below, with the exception of the Unlikely Virtues scale which was embedded in Version 2 of the selection test. The order of presentation of the measures was chosen in order to avoid explicitly priming participants to the possibility that faking can be considered lying. It is possible that once participants were presented with the scales concerning lying they may answer other items differently than if they had not previously thought about faking in terms of lying. Conversely, the *ethic against lying* scale was presented prior to the *beliefs about faking* scale, which measures beliefs that faking is the same as lying, in order to allow participants to think about their ethical beliefs regarding lying generally before asking them whether they believe that faking is the same as lying.

Subjective Norms: Others' Beliefs. Five items were used to assess the extent to which participants believe that significant others in their lives would approve or disapprove of responding desirably on personality-based selection tests (Appendix B). Four of these items were adapted from a scale used by McFarland (2000) and one item was developed specifically for this study. Higher scores on this scale indicate a belief that others think it is acceptable to fake on selection tests.

Subjective Norms: Others' Behavior. Five items were used to examine the extent to which participants believe that others engage in faking on personality-based selection tests. Four of these items were adapted from scales used by Lueke and her colleagues (2001; 2002) and Mueller-Hanson and her colleagues (2003b) and one item was developed specifically for this study (Appendix B). Higher scores on this scale indicate a belief that others engage in faking on selection tests.

Self-efficacy regarding faking. Six items were used to examine participants' selfefficacy for faking (Appendix B). Three of the items were adapted from Wiechmann (2000), two items were adapted from McFarland (2000), and one item was adapted from Mueller-Hanson (2003b). Higher scores on this scale indicate high self-efficacy for successfully faking responses.

Ethic against lying. Seven items were used to measure participants' ethical stance on lying (Appendix B). Four of these items were adapted from a scale used by Christie and Geis (1970) to assess attitudes towards lying in relation to the personality construct of Machiavellianism. Three additional items were constructed specifically for this study. Higher scores on this scale indicate a stronger ethic against lying.

Beliefs about faking. Four items were used to assess participants' beliefs that distorting one's responses on a personality-based selection test is similar to lying (Appendix B). One of these items was adapted from a scale used by Lueke et al. (2001; 2002) and three items were developed specifically for this study. Higher scores on this scale indicate a belief that faking on selection tests is not the same as lying generally.

Knowledge of constructs. A 15-item multiple choice test was developed specifically for this study in order to assess the degree to which participants are aware of

which constructs are being assessed within the test. The items include, in the stem, an item from the personality-based selection test. For each item, participants were instructed to choose the description of the category that the item belonged to, choose which answer was the most desirable by organizational standards, and rate their confidence that they knew the most organizationally desirable response to the item (Appendix C). For each item there is an additional response option of "None of the above." Three items were chosen from each dimension of the selection test to represent a range of obviousness. That is, some items obviously come from a certain dimension (e.g., item number 14), while for other items it is less clear which dimension the item belongs to (e.g., item number 2).

Response distortion. The International Personality Item Pool (IPIP) 17-item unlikely virtues scale ($\alpha = .76$) will be used as the measure of response distortion embedded in Version 2 of the selection test (IPIP, 2001; Appendix D). This scale was constructed to be parallel to the unlikely virtues scale contained in the Multidimensional Personality Questionnaire where it is used as a validity scale (Tellegen, in press).

<u>Procedure</u>

First/Applicant Administration. All applicant testing was completed on-site and was supervised by the organization's hiring personnel. Applicants were instructed to answer honestly, but no warnings regarding faking were provided. There are two normative cutoffs that must be exceeded in order for an applicant to receive an interview. In Version 1 of the selection test, applicants must score at or above the 10th percentile on each dimension and must score at or above the 25th percentile on the overall score, a

summation of standardized dimension scores. Applicants who take Version 2 must score at or above the 11th percentile on each dimension and must score at or above the 23rd percentile on the overall score. These cutoffs were designed to, and do in fact, eliminate approximately 30% of all applicants. Applicants exceeding the minimum cutoffs are interviewed by current managers who make all final hiring decisions. The organization estimates its current selection ratio to be approximately 30%.

Second/Incumbent Administration. Between August and November, 2004, participants were re-administered only the personality-based portion of the selection instrument. This time period was chosen for two reasons: First, it was necessary to allow a sufficient amount of time between the applicant and research administrations of the test in order to prevent participants from simply recalling how they had responded to the items as an applicant. Second, many participants are seasonal workers and are laid off in mid-November. Participants were assured that no one within the organization will ever have access to their individual responses to this administration of the test in both the written consent form and the verbal protocol (Appendix E; F). Participants completed the test during their normal working hours and received their normal hourly wage for participation.

After completion of the personality-based portion of the selection test, all participants were reminded of the confidentiality of their responses and were again requested to respond honestly to the remaining experimental measures. After completion of all measures, participants were administered a second consent form (Appendix G) requesting their permission to obtain their applicant administration test scores and

supervisory performance ratings. Participants had access to debriefing forms once all participants had completed the study (Appendix H).

Supervisory performance ratings were obtained in late November from participants' current supervisors. After agreeing to participate in the current study by signing and dating the consent form (Appendix I), supervisors completed a performance appraisal for each participant.

RESULTS

Construct Validity of the Selection Test

Responses to the incumbent administration of the personality-based test were subjected to an exploratory factor analysis to assess the factor structure of the test. Incumbent responses were factor analyzed, as opposed to applicant responses, due to the findings of previous research indicating higher intercorrelations among scales and the emergence of different factor structures in applicant settings as opposed to research settings (e.g., Barrick & Mount, 1996; Schmit & Ryan, 1993; Weekley et al., 2003). However, the a priori dimensions of the test did not emerge as pure factors. Despite concerns about the low person-to-item ratio (Ford, MacCallum, & Tait, 1986), exploratory factor analyses were conducted to investigate the factor structure of the test. On the basis of a series of rational item groupings and exploratory factor analyses with varimax rotation two correlated factors emerged accounting for 37.25% of the variance in responses. Analysis of the items contained in each factor confirmed that one factor was composed of items similar to items generally used to assess emotional stability and the other factor was composed of items similar to those generally used to measure conscientiousness. All factor loadings for each factor were between .42 and .75.

The emotional stability and conscientiousness scales will be used to test all relevant hypotheses. Additionally, the five a priori, heterogeneous, scales and the overall test score will also be used to test all relevant hypotheses as this represents how the test is used in practice. The emotional stability and conscientiousness scale scores were computed by averaging the item responses such that higher scores indicate more of the construct. The apriori dimension scale scores were computed by summing the empirically coded test responses and standardizing according to previously established

norms in accordance with how these dimensions are used in practice. The overall test score was computed by summing the standardized dimension scores.

Selection Test Difference Scores, Scale Reliability, and Effect Sizes

Difference scores (d-scores) were computed for all applicants by subtracting incumbent scores from applicant scores. This was done for the emotional stability and conscientiousness scales as well as for the test dimensions and overall score. Researchers have argued that d-scores are appropriate when these scores represent a construct of substantive interest, as when one expects a Participant × Treatment interaction (Tisak & Smith, 1994). Other researchers have criticized the use of d-scores because these scores may exhibit low reliability (e.g., Edwards, 2002). However, Rogosa, Brandt, and Zimowski (1982) demonstrate that d-scores do not necessarily exhibit low reliability, and can, in fact, be an accurate and valuable measure of individual change even in situations where the reliability is low. In the current study, d-scores represent a construct that is conceptually meaningful in that it reflects the amount of response inflation occurring as a function of the setting in which test scores were obtained (i.e., applicant and incumbent settings). Thus, a Participant × Treatment interaction was expected because of the assumption, and finding, that some individuals inflate their responses more in an applicant context relative to a research context. Reliability information for the emotional stability, conscientiousness scales, test dimension scales, and social desirability scores, as well as the respective d-scores, are contained in Table 1. The d-score reliabilities were estimated with an equation provided by Rogosa and his colleagues (Table 3, Assumption 0: 1982). Reliability information was not available for the overall test score.

	Applicant	Research			
Variable	Context Rel. ¹	Context Rel. ¹	D-score Rel. ²	Effect Size ³	t4
Emotional Stability	.86	.83	.67	0.83	10.24
Conscientiousness	.77	.76	.58	0.89	10.42
Adaptability	.51	.65	.36	0.79	8.78
Confidence & Friendliness	.43	.76	.44	0.62	7.13
Productivity & Quality	.64	.74	.52	1.01	11.23
Reasoning & Problem Solving	.55	.63	.37	0.21	5.04
Ease of Supervision	.73	.71	.51	0.99	11.19
Overall Test Score				1.14	13.05
Social Desirability	.67	.79	.58	1.34	4.35

Table 1: Paired Samples t Tests, Effect Sizes, and Reliability Information

NOTE: n = 151.

¹Alpha reliability estimates.

²D-score reliability computed with an equation provided Rogosa et al. (1982).

³Effect sizes were computed by subtracting the mean score for the incumbent setting from the mean score for the applicant setting and dividing by the pooled standard deviation.
 ⁴All t-values are significant (p < .01); df = 149

In the applicant context, the alpha reliability of the majority of test dimensions scales is quite low in the current sample, although acceptable levels of reliability were obtained for the emotional stability and conscientiousness scales. Previous analyses on a much larger applicant database indicate that the alpha reliability for all test dimensions is between .65 (Reasoning & Problem Solving) and .82 (Ease of Supervision). It is likely that the alpha reliability estimates underestimate the true reliability of the test dimensions because the items are empirically-keyed, resulting in lower item variance generally, and because of the restricted range of the applicant score distribution. The alpha reliability estimates obtained for the incumbent context were generally at acceptable levels, but the reliability of the Adaptability and Reasoning & Problem Solving scales were slightly below traditionally acceptable levels ($\alpha = .65$ and .63, respectively). The d-scores generally exhibited low reliability which is not unexpected given the substantial positive correlations between the applicant and incumbent test scores and the low reliability estimates observed for the applicant scores.

Paired-sample t-tests and effect size estimates for the difference between the means obtained in the applicant and incumbent settings are also contained in Table 1. Effect sizes were computed by subtracting the mean score for the incumbent setting from the mean score for the applicant setting and dividing by the pooled standard deviation. The positive effect sizes indicate that higher mean scores were obtained in the applicant setting than in the incumbent. All mean differences were significant and all effect sizes were moderate to large.

Descriptive Statistics

Table 2 contains means, standard deviations, reliabilities, and intercorrelations for all variables. There were high to moderate correlations between the selection test scores in both the applicant and the incumbent context, but smaller intercorrelations across the two contexts. Note the substantially lower variance obtained in the applicant as opposed to the incumbent context. In combination with the evidence of higher mean scores, a ceiling effect, in this particular applicant context, this suggests that faking served to depress the variance of scores in this context.

Hypothesis Tests

Social Desirability. Hypothesis 1 states that applicant social desirability scores will have a small to moderate correlation with d-scores. Examination of the correlations between social desirability scores and d-scores (Table 2) indicates that applicant social desirability has a small to moderate relationship with d-scores. However, due to the very small sample size available for these analyses no firm conclusion can be made about the

Table 2: Descriptive Statistics and In	Itercorrela	ations								ļ									
	Mean	SD	٢	2	3	4	5	9	7	8	6	10	11	. 12	3	14 、	5 1	6 1	2
Demographics																			
1. Gender ¹	0.63	0.48																	
2. Age	39.52	16.82	-11																
Applicant Test Scores																			
Emotional Stability	4.54	0.43	8 0.	.07	(98)														
4. Conscientiousness	4.45	0.35	8 0.	.05	49	(77.)													
5. Adaptability	0.59	0.62	.05	.16	¥.	9	.51)												
6. Confidence & Friendliness	0.51	0.53	.05	<u>-</u>	4	.19	.19 (.43)											
 Productivity & Quality 	0.55	0.66	1 2	.18	.35	<u>8</u> .	.56	.29	8										
8. Reasoning & Problem Solving	0.51	0.52	02	<u>6</u>	8 0 [.]	4	.16	.17	14 (.55)									
9. Ease of Supervision	0.48	0.69	.15	<u>ເ</u>	.76	.53	.37	.25	.	.15 (.73)								
10. Overall Test Score	0.69	0.52	.	.12	2	.68	.72	.55	.78	4	.72	I							
11. Social Desirability	4.13	0.36	.05	£.	8	.55	.52	.65	<u>6</u>	. 9	<u>.</u> 63		67)						
12. Cognitive Ability	111.42	13.67	03	<u>6</u>	<u>5</u>	§	<u>.</u>	.05	8	.92	11	53	5	55)					
Incumbent Test Scores																			
13. Emotional Stability	4.14	0.53	28	.05	.53	9	E.	9	.03	8 9.	.33	19	13.	02 (3 3)				
14. Conscientiousness	4.10	0.43	08	-33	9.	45	4	8 0.	32	8	÷	2	15 -	60	50	76)			
15. Adaptability	-0.03	0.92	01	14	.15	.15	4	22	:23	<u>.</u> 07	9 .	32	15 -	<u>5</u>	₹¥.	25 (.	35)		
16. Confidence & Friendliness	0.03	0.96	03	8	<u>.</u> 07	8	<u>.</u> 07	.53	2	.16	<u>8</u>	2	8	6	28 28	1 8	38 (.7	(9	
17. Productivity & Quality	-0.28	0.96	08	.25	60 [.]	.45	.25	.17	.42	.03	.14	.32	21 -	03	24 .	. 68	42	28 (.7	(4)
NOTE: Correlations in bold are signif	ficant (p <	: .05).																	
¹ Gender is coded 0 for female and 1	for male.																		

Table 2 (cont.): Descriptive Statistics	and Inter	correla	tions																
	Mean	SD	٢	2	ო	4	2	9	7	ω	თ	5	11	12	13	14	15	16	17
18. Reasoning & Problem Solving	0.39	0.60	07	8	.07	.16	Ş	.16	.13	.87	.16	.38	8	.87	.16	2	÷.	:23	15
19. Ease of Supervision	-0.31	0.90	8	<u>5</u>	4	2	20	22	.13	9	4	¥.	53.	9	. 68	.32	.47	8	.37
20. Overall Test Score	-0.05	0.76	05	20	.24	30.	હં	9	29	2 9	.25	.46	.27	22	.55	2	.74	69	2.
21. Social Desirability	3.60	0.43	18	. 10	.23	.26	.13	8	.19	<u>.</u>	22	. 18	86	8	.51	.38	38.	.17	4
D-scores																			
22. Emotional Stability	0.40	0.48	8 8.	8	.32	.33	.19	<u>6</u>	.36	 2	32	. 29	ဗ္ဂ	8	63	.13	- 36	- 22	.18
23. Conscientiousness	0.35	0.41	.16	20	ંગ	.38	<u>.</u> 24	8 9.	.35	. 18	¥.	.38	¥.	. 13	-12	-92	.13	-17	2
24. Adaptability	0.62	0.87	.05	03	8 9.	.16	.27	. -	.15	.05	.16	.18	24	<u>5</u>	9	.16	- 76	.35	.27
25. Confidence & Friendliness	0.47	0.82	<u>.</u> 07	05	<u>6</u>	9 .	Ş	100	4	08	4	6 0 [.]	.27	8	-26	.16		8	2
26. Productivity & Quality	0.83	0.91	.17	- 12	.15	5	4	03	. 28	<u>.</u> 07	.18	22	.25	8	- 27	۲	- 28	21	.75
27. Reasoning & Problem Solving	0.12	0.30	6 .	02	8	07	20	\$	02	8	8	8	12	.15	18	.19	.	-19	.25
28. Ease of Supervision	0.79	0.87	12	19	. 18	20	<u>6</u>		.23	<u>5</u>	.36	22	<u>5</u>	<u>5</u>	ź	24	ŧ	Ř	.27
29. Overall Test Score	0.74	0.70	4.	13	14	.18	20	03	.26	8	.26	24	22	8	47	4	- 28	- 29	.53
30. Social Desirability	0.34	0.39	.17	હં	.37	2	.25	<u>9</u> 9.	ଞ	8	.47	.47	- 53	19	5.	.32	<u>ج</u>	- 96.	28
<u>Experimental Measures</u>																			
31. Others' Beliefs	3.00	0.88	03	05	8.	02	07	03	.05	07	13	8 [.]	<u>.</u> 07	60	÷	6 0.	03	5	8
32. Others' Behavior	3.44	0.69	<u>.</u>	36	<u>8</u>	8	17	<u>5</u>	<u>.</u>	<u>.</u>	<u>.</u>	8	6	8	4	20	- 29	.03	3
33. Self-efficacy	3.16	0.83	15	13	8	03	13	05	03	6 0.	03	03	.23	<u>9</u>	8	.15	. 19	- 80	5
34. Ethics Against Lying	3.63	0.70	16	.15	<u>ю</u>	-19	. 19	8	4	8 9	4.	.18	58	9	.15	2	22	8	.25
35. Beliefs About Faking	2.33	0.86	6 0 ⁻	<u>6</u>	8	8	05	<u>.</u>	<u>5</u>	<u>5</u>	60 [.] -	- 02	.07	<u>5</u>	<u>ہ</u>	<u>5</u>	.05 -	8	6
36. Knowledge of Construct	8.29	2.91	<u>.</u> 07	£.	<u>,</u>	2	.26	5	.37	Ş	.19	.38	6 0.	59	8	8	<u>.</u> 33	<u>.08</u>	19
37. Knowledge of Answer	4.12	0.67	6 0.	90	<u>8</u>	.32	<u>8</u>	. 2 8	.38	32	39	.42	<u>.</u> 31	Ł	.20	.1	89.	.17	.15
38. Confidence in Knowledge	4.11	0.64	8 .	8	- 8	.16	9 .	.07	.28	8 9.	8	.17	19	8	8	-01	<u>5</u>	6	8
39. Overall Performance	6.02	1.87	02	90.	06	9 0	06	11	03	09	02	-00	- 30	.01	.02	20	- 02	90	.16
NOTE: Correlations in bold are signifi	icant (p <	.05). /	Npha	relia	bilitie	s are	cont	ained	alon	g the	diag	onal.							

ת $\dot{\mathbf{s}}$ ¹Gender is coded 0 for female and 1 for male.

Table 2 (cont.): Descriptive Statistics	and l	ntero	orrela	tions																
	18	19	20	21	22	23	24	25 ;	26	27 .	28 2	6 6	3	1 3,	3	34	35	36	37	38
18. Reasoning & Problem Solving	(.63)																			
19. Ease of Supervision	.50	(.71)																		
20. Overall Test Score	4 .	.73	ł																	
21. Social Desirability	8	.53	<u>8</u>	.79)																
D-scores																				
22. Emotional Stability	-12	39	4 -	-36 (.67)															
23. Conscientiousness	6 .	15	-30	-17	.42	58)														
24. Adaptability	8	35	57	.31	.52	.31 (36)													
25. Confidence & Friendliness	-17	26	55	-50	8.	.25		4												
26. Productivity & Quality	07		-53	.35	.	. 8 2	39	.33 (52)											
27. Reasoning & Problem Solving	64	23	35	.15	2	.13	.25	8	25 (37)										
28. Ease of Supervision	89	69	-26	.37	99.	4 .	6	89	8	.19 (51)									
29. Overall Test Score	. .	55	- 75	4	.65	.61	.75	. 67	74	ද	78	,								
30. Social Desirability	15	16	29	.52	.52	.56	20	. 59	8	26	2	29 (.)	(8)							
Experimental Measures																				
31. Others' Beliefs	8	- <mark>2</mark>	8.	.15	-07	- 70.	8	33	- 60	2	12	33	32 (.7	(9)						
32. Others' Behavior	<u>5</u>	25	-24	05	-19	2	.19	<u>8</u>	<mark>ک</mark> ۲	8	24	2	5	26 (.7)	6					
33. Self-efficacy	1	07	8 [.]	8	<u>.</u> 07	.13	÷.	8	1 0	19	8	3). 03	5 . 60	7 (.8	2				
34. Ethics Against Lying	8	. 29	.27	20.00	.13 -	8	60	8	1 6	- 8	19 .	16 -	1	231	9 -2	5 (.84	a			
35. Beliefs About Faking	6 9	15	05	8	<u>6</u>	<u>.</u> 05	8	8	8	. 16	 80	20.1	15 .4	47 .3	ູ ເບິ່	, 1 0	1 (.89	~		
36. Knowledge of Construct	.36	.15	24	8.	<u>8</u>	21	.19 -	و	10	15	5 8	- 8C	25(35 .0	30	5.2	5.07	(.67)	_	
37. Knowledge of Answer	4 .	8 .	20	<u>.</u> 07	.25	.23	12	ق	15 -	32.	3.	16	19 -	0.	5.1	6	20.	.38	.92	_
38. Confidence in Knowledge	.13	90. -	5	8	8	.27	<u>.</u> 07	5	17 -	13	80	14	25 .(5	0 [.] 6	1	5	.13	.25	(.92)
30 Overall Berformance	ξ	ç	g	5	ä	ר ג	5	ε	20	16		12	1	ç	5 0 0	5 7	0	5	1	20
	3		3		3		3							?	2	5		3		5

NOTE: Correlations in bold are significant (p < .05). Alpha reliabilities are contained along the diagonal. ¹Gender is coded 0 for female and 1 for male.

relationship between applicant social desirability scores and d-scores. Incumbent social desirability scores also have small to moderate correlations with d-scores, although all relationships are negative, indicating that higher incumbent social desirability scores are associated with lower d-scores.

Hypothesis 3 states that partialling social desirability from applicant scores will not result in a significant change in criterion-related validity. To test this hypothesis performance was first regressed onto each test dimension to obtain an estimate of the validity of the test dimensions. Next, performance was regressed onto social desirability in the first step and regressed onto test dimension in the second step. Examination of Table 3 indicates that, consistent with the hypothesis, partialling social desirability scores from applicant scores did not result in significant changes in the criterion-related validity. However, due to the very small sample size available for these analyses and the absence of applicant test score criterion-related validity no firm conclusion can be made about the effect of partialling social desirability on criterion-related validity.

Construct Validity. Hypothesis 2a states that factor forms will not be significantly different across the two measurement contexts. Hypothesis 2b states that measurement errors of the selection test will be significantly different across the two measurement periods. The previously discussed finding of a lack of factorially pure test dimension scales prevents tests of these hypotheses with those scales. Thus these hypotheses were tested for the emotional stability and conscientiousness scales only. Both emotional stability and conscientiousness were distributed normally in both contexts (i.e., skewness and kurtosis statistics were less than .70). This indicates that it is uneccessary to use asymptotic distribution-free estimation procedures as suggested by previous researchers

Step Variable Etatentry) AR ² Etatentry AR ² Etatentry AR ² Etatentry AR ³ A ³ A ³ A ³ A ³ A ³ A ³ <			Emotional	Ctability,	Conscienti	0000010	ctuchA	, thilida	Confide	nce &
Baseline 194 .003 194 .004 380 .011 Test Dimension 263 .004 194 .003 1565 .090 .014 1 Impression Management -1.665 .090 -1.665 .090 -1.665 .090 2 Test Dimension .382 .006 1.408 .057 963 .075 726 .013 7 Adjusted R ³ .096 (.014) .147 (.069) .165 (.089) .103 (.022 R ² (Adjusted R ³) .095 .001 .322 .008 .064 .001 .033 (.022 Step Variable B (at entry) AR ³ B (at entry) AR	Step	Variable	B (at entry)	ΔR^2	B (at entry)	ΔR^2	B (at entry)		B (at entry)	ΔR^2
1 Impression Management -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .013 2 Test Dimension .382 .006 1.408 .057 963 .075 726 .013 R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 Baseline Productivity & Reasoning & Repension Ease of Quality Focus Quality Focus .006 .013 (.022 Baseline Test Dimension .095 .001 322 .008 .064 .001 330 .022 1 1 .010 322 .008 .066 .010 330 .022 141 .001 265 .090 1 Test	Baseline	Test Dimension	263	1 00.	-194	.003	-194	100	380	.011
² Test Dimension .382 .006 1.408 .057 .963 .075 .726 .013 R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 R ² (Adjusted R ²) .095 (.014) .147 (.069) .165 (.089) .103 (.022 Step Variable B (at entry) AR ² B (at entry) AR ² .001 .330 .022 Test Dimension .095 .001 .322 .008 .064 .001 .330 .022 1 Impression Management .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 .090 .1.655 <td>-</td> <td>Impression Management</td> <td>-1.665</td> <td>060.</td> <td>-1.665</td> <td>060.</td> <td>-1.665</td> <td>060</td> <td>-1.665</td> <td>060</td>	-	Impression Management	-1.665	060.	-1.665	060.	-1.665	060	-1.665	060
R ² (Adjusted R ²) .096 (.014) .147 (.069) .165 (.089) .103 (.022 Reasoning & Productivity & Reasoning & Ease of Quality Focus Productivity & Reasoning & Ease of Quality Focus Overall Test Score Step Variable β (at entry) ΔR ² β (at entry) ΔR ² β (at entry) ΔR ² Impression 095 .001 322 .008 064 .001 330 .022 Impression Management -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 -1.665 .090 .031 .022 .030 .022 .030 .022 .030 .031<	N	Test Dimension	.382	900.	1.408	.057	963	.075	726	.013
Reasoning & Reasoning & Ease of Quality Focus Productivity & Reasoning & Ease of Quality Focus Overall Test Score Step Variable β (at entry) ΔR ²		R ² (Adjusted R ²)		.096 (.014		.147 (.069)		.165 (.08	6	.103 (.022)
Step Variable Quality Focus Problem Solving Supervision Overall Test Score Step Variable β (at entry) ΔR ² β (at entry) Δ (at entry) Δ (at entry)			Productiv	vity &	Reason	ag Se	Ease	of		
Step Variable p (at entry) AR Baseline Test Dimension 095 .001 322 .008 064 .001 330 .022 1 1 Impression Management -1.665 .090 -1.665 .090 -1.665 .090 2 Test Dimension .184 .003 533 .022 141 .001 805 .028 7 rest Dimension .184 .003 1212 .031 131 .021 141 .001 805 .028	đ		Quality F	ocus	Problem (Solving	Superv	ision	Overall Te	st Score
Dasenine Description 095 .001 322 .008 064 .001 330 .022 1<	Step	Variable	B (at entry)	∆R ⁻	β (at entry)	AR-	β (at entry)	AR-	β (at entry)	ΔR ⁺
1 1 Impression Management -1.665 .090 -1.665 .090 2 Test Dimension .184 .003 533 .022 141 .001 805 .028 R ² (Adjusted R ²) .093 (.010) .112 (.031) .091 (.009) .118 (.037)	baseline	Test Dimension	095	.001	322	.008	064	.00	330	.022
2 Test Dimension .184 .003533 .022141 .001805 .028 R ² (Adjusted R ²) .093 (.010) .112 (.031) .091 (.009) .118 (.037	~	Impression Management	-1.665	060.	-1 665	060	-1.665	060	-1.665	060
R ² (Adjusted R ²)	5	Test Dimension	184	.003	533	.022	141	.001	805	.028
		R ² (Adjusted R ²)		.093 (.010		.112 (.031)		.001 (.00	(6	.118 (.037)

ā , Te <u>.0</u> Re 4 ċ

Table 4: Multiple Groups Confirmatory Factor Analyses									
	đ	X²	X²/df	Comparison	∆df	∆X²	TLI	RMSEA	SRMR
M1. Fixed factor pattern, free factor loading,free factor covariance, free factor variance	99	101.41*	1.54	I	I	I	.95	9 0.	02
M2. Fixed factor pattern, equal factor loading, free factor covariance, free factor variance, free factor variance, free error variance	71	102.67*	1.45	M1 vs M2	5	1.26	.95	9 0.	0
M3. Fixed factor pattern, equal factor loading, equal factor covariance, free factor variance, free error	72	105.30*	1.46	M2 vs M3	~	2.63	.95	90.	29
Variance M4. Fixed factor pattern, equal factor loading, equal factor covariance, equal factor variance, free error	74	124.14*	1.68	M3 vs. M4	2	18.84*	.93	<u>.</u> 07	.03
variance M5. Fixed factor pattern, equal factor loading, equal factor covariance, equal factor variance, equal error	81	251.38*	3.10	M4 vs. M5	~	127.24*	.78	.12	Ş
variance <u>NOTE: n = 151. *n < 05</u>									

58

101, p < .Uo.
investigating the effect of faking on measurement invariance (e.g., Smith & Ellingson, 2002). Meredith's (1993) suggested order of model building for examining the equivalence of two tests was used to examine differences in construct validity between the two measurement contexts.

Multiple-groups confirmatory factor analyses (MCFA) with correlated errors were conducted in order to test Hypotheses 2a and 2b. Random parcels composed of two or three items each were used as indicators of the latent factors. Table 4 reports the results of the MCFA described below. Model 1 (M1) tests Hypothesis 2a by constraining only the factor pattern across contexts. This model yielded adequate fit of the model to the data as evidenced by fit indices within the intervals suggested by Hu and Bentler (1999). This indicates that the same item parcels are causing the latent factors across both measurement contexts.

Model 2 (M2) imposed the additional constraint of equal factor loadings. This model also yielded adequate fit of the model to the data. This indicates that the item parcels are equivalent indicators of the latent factors across measurement contexts. Model 3 (M3) imposed the additional constraint of equal covariance. The adequate fit of this model to the data indicates additional support for the construct validity of the scales across measurement contexts.

Model 4 (M4) imposed the constraint of equality of variances across measurement contexts. The significant chi-square difference test indicates that the variance of the latent factors differs across the two contexts. Examination of the variance of the latent factors in the unrestricted model reveals that the variance of the latent factors is greater in the incumbent ($\sigma^2 = .237$ and .169 for emotional stability and conscientiousness,

respectively) than in the applicant context ($\sigma^2 = .182$ and .122 for emotional stability and conscientiousness, respectively). This provides further evidence of a restriction of variance in the applicant as opposed to the incumbent context, although it should be noted that the fit indices for this model do indicate an adequate fit of the data to the model even with this additional constraint.

Model 5 (M5) tests Hypothesis 2b by constraining the measurement errors of the indicators to be equal across contexts. The significant chi-square difference test and the deterioration of the fit indices indicates that measurement errors are not equivalent across the two contexts. Examination of the errors indicates that there is more measurement error associated with the incumbent scores. This is likely due to the greater score variance obtained in incumbent as opposed to the applicant context. Therefore, Hypothesis 2b is supported.

Hypothesis 2c states that the average intercorrelations among the scales will be higher in the applicant than in the incumbent context. Examination of emotional stability and conscientiousness reveals support for this hypothesis. The intercorrelation between these two constructs in the applicant context is .49, and the intercorrelation in the incumbent context is .20. This difference is significant (t (148) = 2.49, p < .05). Thus, Hypothesis 2c is supported.

Criterion-related validity. Hypothesis 4a states that significantly higher criterion-related validity will be observed for the scores obtained in the incumbent as opposed to the applicant context. Table 2 shows that the correlations between applicant test scores and performance ratings are all lower than the correlations between incumbent test scores and performance ratings. T-tests for dependent correlations were used to formally examine

this hypothesis. The conscientiousness scale shows marginally significant differences between the contexts (t (146) = -1.613, p < .10). The differences between the criterionrelated validity of emotional stability is not significant. Therefore, despite that the directions of the differences are consistent with expectations, Hypothesis 4a is not supported. Both the Productivity and Quality dimension and overall test scores exhibited significantly higher correlations in the incumbent as opposed to the applicant context (t (146) = -2.246 and -1.762, p < .05, respectively).

Hypothesis 4b states that the criterion-related validity of the upper portion of the applicant distribution will be lower than for the lower portion of the distribution. A series of hierarchical regressions in which performance was regressed onto applicant test scores in the first step and squared-applicant test scores in the second step were used to test this hypothesis. Table 5 contains the results of the regression analyses. None of the squared predictor terms added significantly to the prediction of performance ratings. Thus, Hypothesis 4b is not supported.

However, the null results could alternatively be explained by the restricted range of applicant test scores available in the current study. For example, if a curvilinear relationship exists across the entire range of applicant test scores such that the relationship is linear at lower test scores (i.e.., applicant test scores not meeting the minimal standards used for selection and not included in the current sample) and either curvilinear or flat at higher test scores (i.e., applicant test scores exceeding the minimal standards included in the current study), then it is possible that a curvilinear relationship still exists even though it was not found with the current restricted data. Analysis of

	ence & liness	ΔR^{2}		.011		.004	.015 (.002)	-		est Score	ΔR^2		.008		.001	(000') 600'	
	Confide Friend	β (at entry)		380		.361	(0			Overall Te	β (at entry)		330		155	0	
quared	ability	ΔR^{2}		.00		.001	.005 (.00	1	e of	/ision	ΔR^{2}		.001		004	.005 (.00	
imensions S	Adapta	β (at entry)		194		109			Ease	Supen	β (at entry)		064		272		
s and Test D	tiousness	ΔR^2		.003		.001	.004 (.000)		ning &	Solving	ΔR^2		.008		.001	(000') 600'	
t Dimensions	Conscient	β (at entry)		.307		.340			Reasol	Problem	β (at entry)	-	322		137		
plicant Tes	Stability	ΔR^{2}		.00		.002	.000 (.000)		vity &	ocus	ΔR^{2}		.001		.002	.003 (.000)	
egressed onto Ap	Emotional	β (at entry)		263		430			Producti	Quality F	β (at entry)		095		.155		
erformance Composite R		Variable		Test Dimension		Test Dimension ²	R^{2} (Adjusted R^{2})				Variable		Test Dimension		Test Dimension ²	R ² (Adjusted R ²) = 140 ^{-*} n < 05	- 140, p / .c.o.
Table 5: Pe		Step	Baseline 1	-	3	•	-				Step	Baseline 1	•	7			

unrestricted applicant data from this organization supports the possibility of this alternative as the standard deviations obtained for applicant test scores in the unrestricted sample are roughly twice that obtained in the current sample. This indicates that the data included in the current sample is substantially restricted.

Selection decisions and rank-order. Hypothesis 5a states that the rank-ordering of test scores will be substantially different in the applicant as opposed to the incumbent context. In order to test this hypothesis, correlations between applicant and incumbent context test scores were computed for the entire distribution. Table 6 contains the results of these analyses for all test dimensions and the overall test scores. With the exception of the Reasoning and Problem Solving dimension, the rank-order correlations were much lower than would be expected when comparing the rank-order of test scores within individuals across time periods. That is, if the same test is given to the same people at two different times, the rank-order correlations between the two time periods would be expected to be quite high (e.g., .70 or greater), which was not found in the current study. Therefore, Hypothesis 5a is supported. However, test-retest unreliability of the test cannot be entirely ruled out as an alternative explanation. Recall that the Reasoning and Problem Solving dimension includes cognitive ability items answered in the applicant context and used to compute scores on this dimension in both contexts so it is not surprising that this rank-order correlation is much higher than the others.

Examination of the rank-order correlations, corrected for range restriction, for the top 50 applicants, selected on the basis of overall applicant test scores, suggests that there were approximately the same amount of changes in rank-order at the top of the distribution as there were for the entire distribution. This is unexpected given previous

Table 6: Rank-order Correlations

			Top 50 Applicant
	Entire	Top 50 Applicant	Scorers -
	Distribution ¹	Scorers	Corrected ²
Emotional Stability	.53	.45	.65
Conscientiousness	.45	.35	.52
Adaptability	.41	.01	.02
Confidence & Friendliness	.53	.43	.51
Productivity & Quality	.42	.32	.51
Reasoning & Problem Solving	.87	.77	.90
Ease of Supervision	.43	.28	.50
Overall Test Score	.46	.24	.56

NOTE: Correlations in bold are significant (p < .05).

 $^{1}n = 151.$

²Values represent correlations corrected for range restriction on applicant test scores.

research findings suggesting that individuals engaging response distortion are more likely to appear at the top of the distribution (e.g., Christiansen & Halaand, 1998).

Hypothesis 5b states that many of the individuals hired on the basis of applicant test scores would not have been hired on the basis of incumbent context scores. Of the 151 individuals in the current sample, all of whom exceeded the minimally acceptable normative score standards on the test as applicants, only 97 participants exceeded these standards based on incumbent test scores. This indicates that over a third of the individuals in this sample would not have been selected in to the organization on the basis of their incumbent test scores and the normative standards used in the applicant context, suggesting that these applicants successfully faked their applicant test scores. Therefore, Hypothesis 5b is supported. It is also possible that test unreliability may have contributed to the obtained results. However, as discussed in the exploratory analyses section below, the 54 applicants that did not meet minimal standards based on incumbent tests scores had significantly greater d-scores than individuals meeting the minimal standards in both contexts. Antecedents. Hypothesis 6a states that perceptions of others' beliefs about the acceptability of faking on selection tests will be positively related to faking. The correlations between other's beliefs and d-scores contained in Table 2 show that perceptions of others' beliefs is significantly related only to Reasoning and Problem Solving d-scores. This correlation indicates that the more an individual perceives that significant others in his or her life believe that it is acceptable to fake on selection tests, the less the individual actually faked his or her scores on the Reasoning and Problem Solving dimension. This relationship is the opposite of that hypothesized. Therefore, Hypothesis 6a is not supported.

Hypothesis 6b states that perceptions of others' behavior in applicant contexts will be related to faking, such that perceiving that other applicants fake on such tests will positively relate to the degree of faking engaged in by the individual. The correlations between others' behavior and d-scores contained in Table 2 show that perceptions of others' behavior is significantly positively related to d-scores for every test dimension except Confidence & Friendliness and Reasoning & Problem Solving. This indicates that the more strongly an individual perceives that other applicants fake on selection tests, the greater the extent of his or her faking in the applicant context. Therefore, Hypothesis 6b is supported.

Hypothesis 7 states that the relationship between ethics against lying and faking will be moderated by beliefs about faking, such that in the presence of a belief that faking is not the same as lying the relationship between ethics and faking will be reduced. Moderated regression analyses, reported in Table 7, reveal no support for this hypothesis.

	2							Confide	nce &
		Emotional :	Stability	Conscienti	ousness	Adaptal	bility	Friendl	ness
Step	Variable	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}
			.020		.005		600 [.]		.003
	Ethics Against Lying	098		028		118		066	
	Beliefs About Faking	022		.022		002		.005	
	2		001		.002		008		.001
	Interaction	022		.024		116		040	
	R ² (Adjusted R ²)		.021 (.000	(000.) 200.		.017 (.000	(.004 (.000)
		Producti	vity &	Reason	s gui	Ease	of		
		Quality F	ocus	Problem (Solving	Supervi	sion	Overall Te	st Score
Step	Variable	β (at entry)	ΔR^2	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}	β (at entry)	ΔR^2
	-		.027		.027		.038		.026
	Ethics Against Lying	217		.010		226 **		161	
	Beliefs About Faking	041		054		.042		013	
	2		.001		.018		004		.007
	Interaction	052		059		-079		060	
	R ² (Adjusted R ²)		.028 (.008	(2)	.045 (.025		.042 (.023	()	.033 (.013)
NOTE	150 *n < 05								

Furthermore, individually, ethics against lying and beliefs about faking are not consistently related to d-scores.

Hypothesis 8 states that self-efficacy for faking will be positively related to dscores. The correlations between self-efficacy and d-scores (see Table 2) reveal no consistent relationships. Self-efficacy is negatively related to d-scores of the Reasoning and Problem Solving dimension, but this is in the opposite direction of that hypothesized. Therefore, Hypothesis 8 is not supported.

Hypothesis 9a states that individuals' knowledge of constructs being assessed will be related to d-scores, and Hypothesis 9b states that self-efficacy will mediate these relationships. Table 2 shows that none of the three operationalizations of knowledge were consistently related to d-scores. Thus, Hypothesis 9a is not supported. The very low power, due to the small sample size (i.e., n = 45), may be partially responsible for the lack of significant relationships. The lack of a significant relationship between selfefficacy and d-scores, as well as between the knowledge measures and d-scores, precludes the possibility of self-efficacy mediating the relationship between knowledge and faking. Thus, Hypothesis 9b is not supported.

Hypothesis 10a states that the relationship between subjective norms, a summative variable of perceptions of others' beliefs and behavior, and faking will be moderated by self-efficacy, such that in the presences of low faking-related self-efficacy the relationship between subjective norms and faking will be reduced. Moderated regression analyses, reported in Table 8, reveal only marginal support for this hypothesis for the emotional stability and conscientiousness dimensions. Thus, Hypothesis 10a is not supported.

	lence &	dliness) Δ R ²	600 [.]			.002		.011 (.000)		est Score		.020			.001		.021 (.000)	
	Confid	Friend	β (at entry)		090	.049		034	(0		Overall T	β (at entry)		600 [.] -	.080		.014	(0	
		bility	ΔR^{2}	.011			000		.011 (.00	of	rision	ΔR^{2}	.049			000		.049 (.03	
		Adapta	β (at entry)		.109	.005		011	(Ease	Super	β (at entry)		047	.162 *		.010	~	
Interaction		ousness	ΔR^{2}	.034			.018		.053 (.033	ing &	Solving	ΔR^{2}	.052 *			000		.052 (.032	
ms, and Their		Conscienti	β (at entry)		.038	.047		.047	(Reason	Problem (β (at entry)		048	033		002	(
ective Nor		Stability	ΔR^{2}	.023			.021		.044 (.024	vity &	ocus	ΔR^{2}	.035			.014		.049 (.029	
Self-efficacy, Subje		Emotional	β (at entry)		.011	.055		.058		Producti	Quality F	β (at entry)		.043	.122		.080		
: D-scores Regressed onto			Variable	1	Self-efficacy	Subjective Norms	7	Interaction	R ² (Adjusted R ²)			Variable	t.	Self-efficacy	Subjective Norms	2	Interaction	R ² (Adjusted R ²)	n = 150; *p < .05.
Table 8			Step									Step							NOTE:

Table 9:	D-scores Regressed onto S	<u>self-efficacy, Ethic</u>	s, Beliefs,	and Their Inte	raction				
								Confide	nce &
		Emotional	Stability	Conscientio	ousness	Adapta	bility	Friendli	ness
Step	Variable	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}
-			.022		.018		.017		.010
	Self-efficacy	.033		.061		660 [.]		089	
	Ethics Against Lying	060 [.] -		014		095		087	
	Beliefs About Faking	030		.007		026		.027	
N	•		.025		.064		.017		.019
	Self-efficacy*Ethics	.003		128 *		.110		.140	
	Self-efficacy*Beliefs	.114		.084		055		.124	
	Ethics*Beliefs	.019		.081		158		026	
ო	~		.021		.001		.002		.014
	Three-way Interaction	114		023		060		.160	
	R ² (Adjusted R ²)		.069 (.023	0	.083 (.038		.035 (.000		.044 (.000)
		Producti	vity &	Reason	a su	Ease	of		
		Quality F	ocus	Problem S	Solving	Supervi	sion	Overall Te	st Score
Step	Variable	β (at entry)	ΔR^2	β (at entry)	ΔR^2	β (at entry)	ΔR^2	B (at entry)	ΔR^2
			.033		.047		.039		.026
	Self-efficacy	.092		055		017		.008	
	Ethics Against Lying	196		003		230 *		159	
	Beliefs About Faking	063		041		.047		015	
0			.027		.028		900		.012
	Self-efficacy*Ethics	186		001		.067		.034	
	Self-efficacy*Beliefs	.110		.046		030.		.066	
	Ethics*Beliefs	.027		043		083		074	
(T)	~		000		.001		.014		000
	Three-way Interaction	.013		.012		169		012	
	R ² (Adjusted R ²)		.060 (.014	(.076 (.031		.059 (.013)		.038 (.000)
NOTE: n	i = 150; *p < .05.								

Hypothesis 10b states that the relationship between ethical beliefs, beliefs about faking, and faking will be moderated by self-efficacy, such that in the presence of lowfaking related self-efficacy the relationship between ethical beliefs and faking will be reduced. Moderated regression analyses, reported in Table 9, reveal no support for this hypothesis.

Exploratory Analyses

Three sets of exploratory analyses were performed to gain a greater understanding of the patterns in the data. Specifically, split-group analyses, moderated regression analyses, and polynomial regression analyses were performed to further examine the relationship between faking and performance.

Previous researchers have examined differences in criterion-related validity for different portions of the applicant distribution to assess the affect of faking on validity (e.g.,Haaland & Christiansen, 1998; Mueller-Hanson et al., 2003). Table 10 presents the split-group correlations of applicant scores, incumbent scores, and d-scores with overall performance ratings and also presents t-tests and effect size estimates for the difference between the means obtained by the two groups. The distribution was dichotomized on the basis of whether or not the individuals' incumbent scores exceeded the minimally acceptable score standards established for the applicant setting. The group failing to meet the minimal score standards as incumbents had significantly lower applicant and incumbent mean scores, but significantly greater d-scores. This pattern suggests that these individuals engaged in greater amounts of faking than the individuals exceeding the minimal standards in both contexts.

Table 10: Split-group Correlations and Effect Sizes

	Pass ¹	Fail ¹	Effect Size ²	t ³
Applicant Test Scores	n=95	n=54		
Emotional Stability	.00	22	37	-2.19
Conscientiousness	.06	01	40	-2.35
Adaptability	03	21	53	-3.10
Confidence & Friendliness	07	27	51	-3.02
Productivity & Quality	04	09	40	-2.38
Reasoning & Problem Solving	01	28	36	-2.12
Ease of Supervision	.03	20	35	-2.08
Overall Test Score	03	35	67	-3.95
Incumbent Test Scores				
Emotional Stability	17	.14	-1.06	-6.25
Conscientiousness	.20	.12	94	-5.56
Adaptability	21	.03	-1.31	-7.70
Confidence & Friendliness	08	17	-1.00	-5.91
Productivity & Quality	.09	.16	-1.29	-7.61
Reasoning & Problem Solving	.03	13	48	-2.85
Ease of Supervision	03	.14	-1.45	-8.56
Overall Test Score	06	.01	-2.07	-12.20
<u>D-scores</u>				
Emotional Stability	.18	33	.77	4.52
Conscientiousness	11	15	.57	3.39
Adaptability	.16	16	.88	5.16
Confidence & Friendliness	.02	.07	.79	4.68
Productivity & Quality	12	24	.96	5.68
Reasoning & Problem Solving	09	22	.28	1.95
Ease of Supervision	.07	27	.88	6.31
Overall Test Score	.03	24	1.12	7.95

NOTE: Values in bold are significant (p < .05).

¹"Pass" refers to those individuals meeting or exceeding the minimal applicant score standards based on incumbent scores, and "Fail" refers to those individuals failing to meet or exceed the minimal applicant score standards based on incumbent scores.

²Effect sizes were computed by subtracting the mean score for the incumbent setting from the mean score for the applicant setting and dividing by the pooled standard deviation; negative values indicate that higher scores were obtained for the group that met or exceeded the minimal applicant score standards based on incumbent scores.

 3 df = 149.

An examination of Table 10 reveals that the correlations between applicant test scores and performance were predominately negative, although not consistently significant due to the small sample size, for the group who failed to meet the minimal standards as incumbents, while the correlations for the group exceeding the standards are essentially zero. A similar pattern emerges when examining the correlations between dscores and performance for these two groups. Although not significant, the pattern suggests that for individuals who, on the basis of incumbent scores, would have failed the selection test the extent of faking is negatively related to performance. Furthermore, this group of individuals had significantly greater d-scores than the group exceeding the minimal standards in both contexts. This provides further support for the notion that some of the people in the current sample were able to successfully fake their way to passing the minimal standards. However, it is important to note that individuals with a true high score on these test dimensions may not have been able to inflate their responses in the applicant setting because these individuals are already at, or very close to, the highest scores possible on the test. If this ceiling effect were not present, it is possible that similar results would have been observed for the "pass" group as were observed for the "fail" group.

Exploratory analyses were performed to investigate whether faking, operationalized as d-scores, moderates criterion-related validity. Hypotheses 4a and 4b also examined related, but different, phenomenon as the current analyses. Hypothesis 4a examined the differences in criterion-related validity for scores obtained in the two contexts. In contrast, the moderated regression analyses described here test whether the validity of applicant test scores changes over different levels of faking. Hypothesis 4b examined the differences in validity for the top portion of the distribution, where faked scores are most likely to reside, and the lower portion of the distribution, likely to have less faked scores. The current moderated regression analyses are less influenced by the restricted range of applicant test scores that may have resulted in the null result for Hypothesis 4b.

Table 11 shows that the interaction of applicant scores and d-scores was significant for emotional stability, but not for conscientiousness. Figure 2 contains a graph of the significant interaction. For emotional stability, faking moderates the relationship between applicant scores and performance such that in the presence of a relatively high degree of response inflation, test scores relate negatively to performance. In the presence of relatively low amounts of faking, test scores relate positively to performance. A similar pattern was also observed for Ease of Supervision and overall test scores. It is important to note that the results of these analyses are purely exploratory and cannot be interpreted as strong support for the notion that faking moderates the criterion-related validity of applicant test scores. This is especially true when one considers the compounding of unreliability inherent in d-scores and the interaction term.

As discussed previously, some researchers argue that polynomial regression using the d-score components is a more appropriate analytical technique than analyses involving d-scores (cf. Edwards, 2002). Thus, the interactive effects of applicant and incumbent test scores on performance ratings were examined using polynomial regression. As suggested by Edwards (2002), applicant and incumbent test scores were entered in the first step, applicant and incumbent squared scores were entered in the second step, and a multiplicative interaction term for applicant and incumbent scores was entered in the third step of a hierarchical regression predicting performance ratings.

The results, reported in Table 12, show that the interaction of applicant and *incumbent test scores was significant for emotional stability*, Ease of Supervision, and **overall test scores**. Figure 3 shows the nature of this interaction for emotional stability. The graphs of the interactions for Ease of Supervision and overall test scores are very

Table 11	1: Performance Regressed onto	Applicant Test Di	imension S	scores, D-scor	es, and Th	eir Interaction	_		
		- Incontraction	Ctobility.				, Hild	Confide	nce &
						o /ot onto /o			
Step	Variable	p (at entry)	AR	p (at entry)	24	p (at entry)	27	p (at entry)	YA
	1		.007		.039		6 0.		.011
	Applicant Test Dimension	175		.720		188		380	
	D-score	240		918		014		.002	
	2		• 030		.015		.013		001
	Interaction	-1.549 *		-1.435		404		146	
	R ² (Adjusted R ²)		.037 (.017		.054 (.035)		.018 (.000		.013 (.000)
		Producti	vity &	Reason	ng &	Ease	of		
		Quality F	ocus	Problem \$	Solving	Superv	ision	Overall Te:	st Score
Step	Variable	β (at entry)	ΔR^{2}	β (at entry)	ΔR^2	β (at entry)	ΔR^{2}	β (at entry)	ΔR^{2}
	-		.039		.033		.015		.021
	Applicant Test Dimension	.065		-1.300 *		.060		230	
	D-score	416 *		.979		279		310	
	2		.003		.003		.028 *		.027 +
	Interaction	171		-1.757		540 *		803 *	
	R ² (Adjusted R ²)		.042 (.022	(.039 (.005)		.043 (.023)		.048 (.028)
NOTE:	n = 149; *p < .05.								



similar to the graph for emotional stability and thus are not reproduced here. Examination of Figure 3 suggests that regardless of the direction of the difference between applicant and incumbent scores, greater differences predict lower performance. This indicates that, for emotional stability, Ease of Supervision, and overall test scores, the absolute difference between applicant and incumbent scores relates negatively to performance ratings. Contrary to theory and expectations, this suggests that response inflation as well as response deflation influences the criterion-related validity of applicant test scores.

	ence &	liness	ΔR^2	.011			.005			<u>.004</u>		.020 (.000)		est Score	ΔR ²	.021			.001			.040		.062 (.029)	
nteraction	Confide	Friend	β (at entry)		378	002		.416	047		.328	()		Overall Te	β (at entry)		540	.310		197	.058		1.405 *	()	
and Their Ir		ability	ΔR^{2}	<u>,004</u>			.022			.029		.055 (.022	e of	/ision	ΔR²	.015			900 [.]			.040		.061 (.028	
sion Scores a		Adapta	β (at entry)		202	.014		097	249		.741		Ease	Super	β (at entry)	1	219	.279		292	-004		.948 *		
Test Dimens		ousness	ΔR^{2}	• 039			.003			.025		.067 (.034)	ing &	Solving	ΔR²	.033			.003			.003		.039 (.005)	
d Incumbent		Conscienti	β (at entry)		200	.922 *		.367	.279		2.446		Reasor	Problem	β (at entry)		-1.300 *	616.		390	.220		-1.757		
oplicant and		Stability	ΔR^{2}	.007			.013			690 [.]		.089 (.057	vity &	-ocus	ΔR²	• 039			.007			.002		.048 (.015	
imance onto A		Emotional	β (at entry)		417	.242		147	574		2.964 *		Producti	Quality F	β (at entry)		352	416 *		.163	.092		.201		
Polynomial Regression of Perfo			Variable		Applicant Test Dimension	Incumbent Test Dimension		Applicant Test Dimension ²	Incumbent Test Dimension ²		Interaction	R ² (Adjusted R ²)			Variable		Applicant Test Dimension	Incumbent Test Dimension		Applicant Test Dimension ²	Incumbent Test Dimension ²		Interaction	R ² (Adjusted R ²)	= 149; *p < .05.
Table 12:			Step	-			2			e					Step	-			7			с С			NOTE: n



DISCUSSION

The results of this study indicate that individuals, on average, respond to personality-based self-report tests more desirably in applicant as opposed to incumbent settings. Furthermore, the results suggest that different individuals may inflate their responses to different degrees in applicant settings and this inflation may affect rankordering as well as selection decisions. The low power confirmatory factor analyses suggest that applicant faking did not erode the construct validity of the personality-based measures used in the current study as both factor patterns and loadings appear to be equal across applicant and incumbent contexts. However, the significantly higher intercorrelation obtained in the applicant context for emotional stability and conscientiousness suggests the opposite conclusion with regards to construct validity.

Some of the analyses, such as the d-score moderation analyses, suggest that the criterion-related validity of applicant test scores may be moderated by applicant faking. However, other analyses, such as the examination of different validity for applicant and incumbent test scores and the analysis of a curvilinear relationship beween applicant scores and performance, suggest that criterion-related validity is not attenuated or moderated by applicant faking. Finally, the degree of faking engaged in by an individual is positively related to the degree to which an individual endorses a perception that others engage in faking in applicant settings. Table 13 formally summarizes the hypothesizes and whether each one was supported by the data. The remainder of this paper will address five central questions in the faking literature and how the results of this study contribute to this knowledge base.

Table 13: Hypothesis Summary

Hypothesis	Result
Hypothesis 1: Scores on a social desirability scale will have a small to moderate correlation with difference scores.	Supported
Hypothesis 2a: Factor forms will not be significantly different across the two measurement periods (i.e., applicant administration and research administration), that is there will be configural invariance of the personality factors across both	
administration periods.	Supported
Hypothesis 2b: Measurement errors of the personality test will be significantly different across the two measurement periods.	Supported
<u>Hypothesis 2c</u> : Average intercorrelations among the scales will be higher when the inventory is administered for application purposes than when it is administered for research purposes.	Supported
Hypothesis 3: Partialling social desirability scale scores from applicant personality scores will not result in a significant change in criterion-related validity.	Not supported
<u>Hypothesis 4a</u> : Criterion-related validity will be significantly greater for personality scores obtained in a non-motivating context (i.e., for research purposes) than scores obtained in a motivating context (i.e., for application purposes).	Not supported
<u>Hypothesis 4b</u> : Criterion-related validity of the upper half of the distribution of applicant sample scores will be less than the criterion-related validity of the bottom half of the distribution.	Not supported
<u>Hypothesis 5a</u> : The rank-ordering of people on the basis of responses obtained in a non-motivating context (i.e., for research purposes) will be substantially different than the rank-ordering of responses obtained in a motivating context (i.e., for application purposes).	Supported
<u>Hypothesis 5b</u> : Some of the people hired on the basis of their responses in a motivating context (i.e., applicant setting) would not have been hired on the basis of their responses in a non-motivating context (i.e., research setting).	Supported
Hypothesis 6a: Perceptions that others believe faking to be an acceptable practice will be related to faking.	Not supported
Hypothesis 6b: Perceptions that others engage in faking will be related to faking.	Supported
<u>Hypothesis 7</u> : The relationship between a self-reported ethic against lying and faking will be moderated by an individual's belief that faking on a selection test is the same as lying, such that in the presence of a belief that faking is not the same as lying the relationship between reporting an ethic against lying and faking will be reduced.	Not supported
Hypothesis 8: Individuals with high self-efficacy for enhancing their responses to a selection test in a desirable way will be more likely to engage in faking.	Not supported

Table 13 (cont.): Hypothesis Summary

Hypothesis	Result
Hypothesis 9a: Knowledge of the constructs being assessed will be related to faking.	Not supported
Hypothesis 9b: The relationship between knowledge of the constructs being measured and faking will be partially mediated by self-efficacy.	Not supported
Hypothesis 10a: The relationship between subjective norms and faking will be moderated by self-efficacy, such that in the presence of low faking-related self- efficacy the relationship between subjective norms and faking will be reduced.	Not supported
Hypothesis 10b: The relationship between ethical beliefs and faking will be moderated by self-efficacy, such that in the presence of low faking-related self- efficacy the relationship between ethical beliefs and faking will be reduced.	Not supported

Does Faking Affect Selection Decisions?

Past research suggests that individuals engage in response inflation by showing that applicants generally score higher than incumbents (e.g., Rosse et al., 1998), applicants score higher than research participants (e.g., Birkeland et al., 2003), individuals incentivized to perform well on personality tests get higher scores than when they are not incentivized (e.g., Mueller-Hanson et al., 2003a), and by showing that when instructed to do so, individuals can inflate their responses (e.g., McFarland & Ryan, 2000). All of these types of designs have unique limitations that open their results to criticism. For example, research that shows applicants tend to score higher than incumbents does not indicate that only some applicants are faking, but rather could indicate that all applicants are uniformly increasing their scores in response to the situation. If true, this would result in little or no change in rank-orders and, thus, no out of order decisions or erosion of criterion-related validity as a consequence of faking. However, the within-subjects design of the current study as well as the use of actual applicants confirms that many individuals do in fact inflate their responses in applicant settings and, furthermore, that individuals engage in different degrees of faking resulting in changes in rank order and even changes in which individuals should be selected.

Echoing the concerns of Mueller-Hanson and her colleagues (2003a), these results suggest personality test scores should not be used in a top-down, or select-in, manner. However, the current results also cast some doubt on the use of these tests for select-out purposes as well. Specifically, the expected gains from implementing such a test may not be realized if a substantial number of people exceed the normative cutoff standards due to response inflation.

The results showing that many of the individuals in this study would not have been hired on the basis of their incumbent test scores suggests the possibility that individuals can fake their way to passing normative cutoff standards. However, as noted previously, the unreliability of the test may have also contributed to this finding. Future research could help to untangle the effects of unreliability vis-à-vis faking within the current sample by obtaining a test-retest reliability estimate that would allow for an assessment of how many people would likely not have passed on the second administration simply due to error. Such research could be performed with a convenience sample of students tested at two time periods. Future research with similar designs and samples using tests with high levels of test-retest reliability would also help to elucidate whether some applicants do in fact fake their way into being selected.

Assuming that applicant faking was largely responsible for the current finding that some people who met the minimal test standards on the basis of applicant but not incombent scores suggests that by not controlling faking on such tests, organizations may be unfairly rewarding those who are inclined to dishonestly raise their scores. This

phenomenon takes on even more importance when organizations use these tests in a topdown selection manner. Perhaps one of the reasons individuals tend to view the use of personality and other non-verifiable self-report tests as less fair than more objective selection procedures such as cognitive ability tests or interviews (Hausknecht, Day, & Thomas, 2004) is because of their awareness that people can effectively lie on such tests to increase their chances of selection. Supporting this notion is recent research showing that individuals with relatively lower fairness perceptions of such tests report engaging in greater levels of faking (McFarland, 2002). Future research using within-subjects designs and field samples should include measures of fairness perceptions to assess this possibility.

Are Social Desirability Scales a Valid Operationalization of Faking?

One of the key goals of the current study was to compare multiple operationalizations of faking in order to contribute to understanding whether social desirability scales are valid measures of faking in applicant contexts and could thus be used to identify faked applicant test scores. The relatively low intercorrelations between applicant social desirability scores and d-scores suggests the possibility that these two operationalizations of faking are unique constructs and do not share substantial variance. Partialling social desirability scores from applicant test scores did result in slight, but insignificant, increases in criterion-related validity for the conscientiousness, Adaptability, and Reasoning and Problem Solving scales. Unfortunately, however, the very small sample size available for these analyses precludes drawing any firm

conclusions from these results.

The prevalence with which social desirability scales are used in practice to identify potentially faked responses (Goffin & Ghristiansen, 2003) indicates a distinct need for researchers to determine whether these scales are actually measuring applicant faking. Research designs using actual applicants and a within-subjects design, such as used by the current study, with larger sample sizes would be helpful in determing whether faking is adequately operationalized by these scales. However, it is important to note that recent simulation research by Schmitt and Oswald (in press) suggests that identifying fakers with social desirability scales and removing those identified from the selection process is likely to have very little positive impact in terms of mean performance of those hired.

Does Faking Affect Construct Validity?

Consistent with the majority of prior research (e.g., Smith & Ellingson, 2002; Weekley et al., 2003), the current study found that applicant and non-applicant test responses result in similar factor forms but dissimilar measurement errors. The current study also found that factor loadings were similar across contexts, a minimum condition suggested to be necessary to conclude measurement invariance across groups (cf., Rock et al., 1978). The results of the current study are likely to be a more veridical reflection of reality than some of the prior studies for two main reasons.

First, many previous studies assessed measurement invariance by comparing an applicant group to a separate research or student group, raising the possibility that the differences found between the two groups were a result of differences inherent in the par-ticular samples and not due to the differing contexts. The within-subjects design of the current study allows the inference that if differences were found between the two measurement contexts, the differences must be due to the context and not to underlying differences between the samples. Second, the current study's use of a field sample allows assessment of measurement differences as a function of real-life applicant motivation instead of artificial laboratory-induced motivation present in many previous studies. Thus, the results suggest that acceptable degrees of measurement invariance are present between applicant and incumbent settings despite significant score differences between the applicant and incumbent contexts. Note, that evidence was found that the interrelationships between constructs was greater in the applicant as opposed to incumbent settings suggesting some erosion of discriminant validity, but this relationship did not appear affect the factor structure of the constructs.

Despite the advantages of the current design over previous research designs examining construct validity, it is important to note that the test used in this study did not exhibit the expected factor structure in either context and thus the results are based on a post hoc factor structure established with the incumbent test responses. Furthermore, the MCFA is generally a low power test which may have prevented the emergence of significant differences between the two contexts. Also, as noted previously, the results indicated that the discriminant validity of the tests were somewhat compromised in the applicant context. Finally the existence of individual differences in the degree of response inflation found between the two contexts, despite the lack of significant factor loading differences observed with the MCFA. Additionally, it is important to note

that some researchers insist that similar error structures must also be observed to conclude factorial invariance (Meredith, 1993).

Given these limitations and conflicting results, it is necessary for future research to replicate these results before any application of the current results can be justified. For example, research using similar designs and tests with more robust and established factor structures would likely provide more generalizable evidence. Additional research comparing alternative measures of personality constructs assessed via self-report instruments would also help to further understanding of the effect of applicant faking on the construct validity of personality-based self-report tests.

Does Faking Affect Criterion-related Validity?

The current study was designed to address the effect of faking on criterion-related validity with a unique research design that has many advantages over previous investigations. However, despite prior evidence of the validity of the selection test used in this study, in the current sample, the test demonstrated very low validity in both the applicant and the incumbent context. Therefore the current study's assessment of the affect of faking on criterion-related validity is exploratory and should not be interpreted as proof that faking either does or does not impact the criterion-related validity of these types of tests.

The results of this study provide conflicting evidence of the effects of faking on criterion-related validity. The pattern of results for some analyses appear to be consistent with laboratory studies of faking, demonstrating that faking can impact criterion-related validity (e.g., Dunnette et al., 1962; Mueller-Hanson et al., 2003a). However, the pattern

of results for some analyses also suggests that faking does not affect criterion-related validity.

Generally, the criterion-related validity of the incumbent scores was greater, or less negative, than for the applicant scores. While these results did not reach traditional significance levels, the pattern suggests that faking may impact criterion-related validity. Additionally, examination of the performance correlations for the group of individuals whose incumbent scores did not exceed minimal selection standards reveals a similar, but stronger pattern. Specifically, the applicant score correlations with performance for this group are all negative, while the incumbent score correlations are predominately positive. Despite this pattern, the majority of these analyses were not significant and thus the results may be due to the generally low validity of the test, test unreliability, or range restriction.

Investigation of a curvilinear relationship between applicant test scores and performance revealed no such relationship, in contrast to the findings of previsous researchers (e.g., Haaland & Christiansen, 1998). However, the restriction of applicant test score range inherent in the current sample may have masked such a relationship because the lower portion of the score distribution, expected to have a positive linear relationship with performance, is missing from the analyses. Future research should investigate this question by using a design similar to this study, but with an unrestricted sample. For example, a similarly designed study performed in the context of a predictive validation study in which the selection test is not used for selection would provide the neccessary data to adequately investigate this question.

D-scores show a pattern of negative relationships with performance for the entire sample, although these relationships are largely non-significant. While the pattern suggests that the greater the amount of response inflation engaged in by an individual the lower his or her performance ratings, the results were nevertheless nonsignificant and potentially due to test unreliability. One piece of evidence suggests the possibility that the results were not solely due to unreliability. If the negative relationships between dscores and performance were due to unreliability then there is no reason to expect that a different pattern would emerge for individuals who did or did not meet minimal score standards on the basis of incumbent test scores. However, the data suggests that the relationship between difference scores and performance is stronger for individuals who did not exceed the minimal selection standards on the basis of their incumbent test scores. Recall that these individuals also had significantly greater levels of faking compared to those individuals exceeding the minimal standards at both time periods. Therefore, focusing solely on traditional significance leads to the conclusion that d-scores are unrelated to faking, but focusing on the pattern of results suggests the opposite conclusion.

D-scores interacted with applicant test scores for the emotional stability and Ease of Supervision scales, as well as for overall test scores. Specifically, individuals engaging in greater levels of faking exhibited a negative relationship between applicant test scores and performance ratings. Individuals engaging in little or no faking had a positive relationship between applicant test scores and performance. This suggests that faking may moderate the relationship between applicant test scores and performance.

However, given the extremely low reliability of both applicant scores and d-scores, the results should be interpreted cautiously.

The polynomial regression analyses suggest that response inflation and response deflation equally negatively impact predicted performance. The results of the polynomial analyses should be interpreted with caution. First, these results could be due to regression to the mean effects in which people with extremely high scores the first time received somewhat lower scores the second time due to error. Second, there is absolutely no theoretical justification for why response deflation in applicant contexts should relate negatively to performance. One could speculate that very unstable people may respond lower at one time than another and that this instability could manifest itself in behaviors on the job. However, such speculation based on the analyses from one sample commits the sin of "letting the empirical tail wag the theoretical dog," (Bedeian & Day, 1994). Other researchers also argue that theory should serve as a precondition for selecting the types of analyses used to test research hypotheses (cf. Schoorman, Bobko, & Rentsch, 1991) and Tisak and Smith (1994) specifically caution against accepting unconstrained polynomial regression models simply because they fit the data better in a particular sample even though no theory exists to explain the results. Thus, before accepting these results as an accurate reflection of reality additional theoretical work and replication of these results is necessary.

Thus, while the results are conflicting, the pattern of results suggests the possibility that the criterion-related validity of personality-based selection tests may be affected by applicant faking. However, future research is necessary in order to examine whether the pattern of results found in the current study are replicable and generalizable.

Future research should specifically seek to examine these relationships, in the context of within-subjects designs using field samples of actual applicants, similar to the current study. However, future research should use different personality-based tests and also should attempt to utilize samples that are not restricted due to selection on the basis of the tests being examined.

There is also the possibility, as suggested by recent simulation research (Schmitt & Oswald, in press), that the validity of personality-based selection tests is so small that faking can only have a marginal impact upon their validity. In some ways, the results of the current study are more supportive of this notion than that faking does or does not impact criterion-related validity. This suggests that research efforts aimed at increasing the validity of these types of tests may yield a larger payoff in terms of predicting job performance than research that continues to investigate faking.

Do Individuals' Perceptions & Beliefs Relate to Faking?

The current study investigated the influence of perceptions that others think it is acceptable to engage in faking, perceptions that others actually fake, ethics against lying, beliefs that faking is the same as lying, self-efficacy, and knowledge of measured constructs. The current study found that only perceptions that others actually engage in faking related to the extent to which an individual engaged in faking.

The finding of a relationship between an individual's perceptions that others engage in faking and d-scores is in need of replication, but is nevertheless interesting. This is interesting because it represents another possible route through which individuals may be persuaded not to fake in applicant contexts. For example, in addition to

traditional warnings that focus on injuctive norms (i.e., what is approved or disapproved), test administrators could make appeals focusing on descriptive norms (i.e., what is commonly done) of stating that most people do not engage in faking on such tests. Prior research has demonstrated that both types of normative appeals contribute to intentions and behavior (e.g., Cialdini et al., in press). Future research would be useful in determining if the addition of disjunctive norms to the injunctive warnings against faking commonly provided to applicants are helpful in further reducing applicant faking.

There are a number of potential reasons for the why the other antecedent measures did not relate to faking. For example, as described previously, d-scores tend to, and did in the current study, have low reliability which may have attenuated the correlations between the antecedent measures and d-scores. Additionally, while the antecedent measures used in the current analysis included some items from prior research with established construct validity, some of the items were developed specifically for this study and thus may not be adequately tapping the intended constructs. It is also possible that the current sample, in addition to being restricted in terms of applicant test scores, is restricted in terms of the antecedent measures as well which could explain why some of the antecedents consistently related to d-scores in the expected direction but did not reach statistical significance (e.g., self-efficacy). A useful follow-up study would be to administer these antecedent measures to an additional sample to determine if in fact the range of responses obtained in the current study is restricted. Another possibility is that participants were engaging in socially desirable responding on the antecedent measures due to the presence of the experimenter and the sensitive nature of some of the questions.

Utilization of a randomized response technique methodology would help to control for this possibility in future research (Fox & Tracy, 1986).

In relation to the knowledge of constructs measure, it is difficult to say whether "correctly" choosing the construct represented by an item is a sufficient measure of individuals' knowledge of the constructs given that the factor analyses failed to support the a prior constructual dimensionality of the test. Additionally, the sample size was quite small for the knowledge measures and thus the associated statistical tests were woefully underpowered. Future research should investigate the usefulness of similar knowledge measures using a more construct valid test and a larger sample size.

Despite the lack of findings for many of the individual beliefs and perceptions included in this study, it is important to continue research to uncover what leads individuals to engage in faking. This is particularly important in light of recent simulation research showing that variability in applicant faking had a larger effect on validity coefficients than either the average magnitude of faking or the total proportion of faked scores in a sample (Komar, Theakston, Brown, & Robie, 2005). Thus, future research should continue to investigate the perceptions and beliefs examined here as well as other individual differences that may influence faking. Verbal protocol analysis of individuals engaging in responding to personality-based tests under both motivating and non-motivating conditions may be particularly helpful in elucidating the process through which individuals determine how to respond to such questions in different contexts. The investigation of person-situation interactions would also further understanding of what conditions, both individual and contextual, motivate individuals to engage in faking and

whether certain situations are more likely to motivate individuals with given beliefs or characteristics more than other situations.

Limitations

Several study limitations exist. First, the sample was severely restricted as only hired individuals were able to participate. Future research could avoid this limitation by examining faking in the context of a true predictive validation study where selection is not based on test scores. The restriction of range is likely the reason that the applicant test scores exhibited non-significant, and sometimes negative, validities with performance ratings for the entire distribution. However, the finding of non-significant predictive validity for the applicant test also represents a limitation of the current study. It should be noted that the validity obtained for this test in the initial concurrent validation study was substantial, with average correlations between test scores and performance ratings of .28. It may also be argued that the use of d-scores in this study represents another limitation. However, as noted previously, d-scores have distinct advantages over other operationalizations of faking (i.e., d-scores represent a direct measure of the amount of inflation occurring as a function of the setting), are conceptually meaningful in this context, and are not inherently flawed (Rogosa et al., 1982).

Conclusion

The results of this study contribute to the literature on faking in a number of ways. First, the large effect sizes found between the applicant and incumbent administrations of

the test indicate that applicants *do* engage in response inflation. Second, the finding that 36% of the current sample did not exceed the normatively established cut scores, on the basis of their incumbent test responses, corroborates previous findings performed in laboratory settings or using social desirability scores to identify faked responses, although the unreliability of the test used in the current study indicates that replication of this finding with a more reliable test is desirable. This study also presents some evidence that faking may impact criterion-related validity. Finally, this study provides evidence that an individual's perceptions of others' behavior in applicant settings relates to individual response inflation. However, the current study failed to find relationships between the other antecedents measured and faking behavior. Future research with similar designs that addresses the limitations of the current study will be useful in determining whether the current pattern of results is replicable and generalizable.
APPENDICES

Entry-Level Selection Research Project Performance Appraisal Form

Your Name:

Employee Name:

		_						_			
Outstanding	Top 10%	0	0	0	0	0	0	0	0	0	0
Excellent	Next 20%	0	0	0	0	0	0	0	0	0	0
Good	Middle 40%	0	0	0	0	0	0	0	0	0	0
Fair	Next 20%	0	0	0	0	0	0	0	0	0	0
Weak	Bot. 10%	0	0	0	0	0	0	0	0	0	0
In the right to evaluate this employee on each of the following log, simply place and "or checkmark in the appropriate circle to mance factor. Use the percendages associated with each rating diding on the appropriate rating (e.g., is this employee in the to ouve seen?). Remember: all of your ratings are completely	Description	Norks well with others. Seems to enjoy being part of a team. Respects coworkers.	Willing to take direction. Open to learning new things from supervisor.	Confident and positive attitude. Friendly toward customers and other employees.	Concerned with the quality of what is produced. Makes certain products and services meet the needs of customers.	Appropriately applies rules and procedures in making decisions. nnovatively solves problems. Identifies process improvements.	Assumes responsibility to ensure the needs of the workplace are satisfied. Takes initiative in making sure things get done.	Completes assignments accurately and efficiently. Maintains a high level of energy and activity throughout the work day.	Flexibly adapts to changing job demands and to changes in the workplace. Maintains composure even in stressful situations.	Works with supervisor to try to make schedule work and help cover difficult times. Shows up up on time when scheduled.	Considering all of the above, plus other factors that may be important, how would you rate this employee overall?
Please use the rating factors. To make a ra the right of each perfo level to guide you in de 10% of all employees confidential.	Performance Factor	Working with Others	Ease of Supervision	Confidence and Friendliness	Quality and Service Focus	Reasoning and Problem Solving	Responsibility and Ownership	Energy and Productivity	Adaptability	Attendance	Overall Performance

APPENDIX A

APPENDIX B

INSTRUCTIONS: Please show how strongly you agree or disagree with each of the following statements by putting a circle around the appropriate number to the right of each item. You may choose not answer any question,

PLEASE NOTE:

- "Answering in a desirable manner," includes slightly exaggerating your responses to make yourself look good, responding to a question in a desirable way even if you do not think that it is completely true, and completely making up answers without thinking about whether the answers are true or not.
- "Selection test" refers to any test that is similar to the one you took as an applicant.

Pleas	e circle the appropriate number to the right of each item to show strongly you agree or disagree with it.	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
	Subjective Norms: Others' Beliefs					
1	If I answered in a desirable manner on a selection test, most of the people who are important to me would disapprove.*	1	2	3	4	5
2	No one who is important to me thinks it is OK to answer in a desirable manner on a selection test.*	1	2	3	4	5
3	Most people who are important to me will look down on me if I answer in a desirable manner on a selection test.*	1	2	3	4	5
4	My parents would approve of me answering in a desirable manner on a selection test.	1	2	3	4	5
5	My friends and family would disapprove of me enhancing my responses on a selection test in order to make a good impression.*	1	2	3	4	5
	Subjective Norms: Others' Behavior					
6	Other people probably answer in a desirable manner on selection tests in order to get a better score.	1	2	3	4	5
7	Most applicants would not hesitate to answer in a desirable manner on a selection test.	1	2	3	4	5
8	In most hiring situations, applicants do NOT distort or enhance their responses to selection tests.*	1	2	3	4	5
9	Everyone changes their answers on selection tests to appear more desirable.	1	2	3	4	5
10	Most applicants exaggerate their answers to selection tests in order to make a good impression.	1	2	3	4	5

Pleas	se circle the appropriate number to the right of each item to show strongly you agree or disagree with it.	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
	Self-efficacy					
11	I am confident that I could receive a higher score on selection tests by exaggerating my responses.	1	2	3	4	5
12	I'm confident I could figure out how to get a higher score on selection tests.	1	2	3	4	5
13	It would be easy for me to increase my score on a selection test by answering in a desirable way.	1	2	3	4	5
14	If I want to, I could increase my score on a selection test.	1	2	3	4	5
15	I could make myself look better on a selection test by responding dishonestly.	1	2	3	4	5
16	I could respond in a very desirable way to selection tests, if I chose to.	1	2	3	4	5
	Ethic Against Lying					
17	People should never lie.	1	2	3	4	5
18	I never tell lies to other people.	1	2	3	4	5
19	I think it is sometimes necessary and ethically acceptable to lie to other people.*	1	2	3	4	5
20	The best way to handle people is to tell them what they want to hear even if it is not true.*	1	2	3	4	5
21	Honesty is the best policy in all cases.	1	2	3	4	5
_ 22	There is no excuse for lying to someone else.	1	2	3	4	5
23	All in all, it is better to be honest than to be dishonest.	1	2	3	4	5
	Beliefs About Faking					
24	I believe that answering in a desirable manner on a selection test is not the same as lying.	1	2	3	4	5
25	Lying is different than exaggerating responses on a selection test.	1	2	3	4	5
26	Making yourself look good on a selection test by answering in a desirable manner is different than lying.	1	2	3	4	5
27	Exaggerating responses on a selection test in order to make a good impression is the same as lying.*	1	2	3	4	5

* denotes reverse keyed items
 NOTE: All scale labels will be omitted in the participant version of these measures.

APPENDIX C

Knowledge of Constructs

INSTRUCTIONS: Each question below presents one of the items that you responded to in the previous phase of the experiment. For each item: A) fill in the circle for the category that best describes the item; 2) fill in the circle that indicates how you think the organization would like you to respond to the item; 3) fill in the circle that indicates how confident or sure you are that you know how the organization would like you to respond to this item.

1. "I take time out for others."		
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
O ² How confident and friendly a person is. O 3 How productive and focused	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.		
2. "I usually arrive early for appointn	nents."	
2. "I usually arrive early for appointmA) Which category best describes this item?	nents." B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. 	B) How would the organization like you to answer this item?	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused 	B) How would the organization like you to answer this item? O 1 Strongly Disagree O 2 Disagree	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. 	B) How would the organization like you to answer this item? 0 1 Strongly Disagree 0 2 Disagree 0 3 Neither	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. 	B) How would the organization like you to answer this item? 0 1 Strongly Disagree 0 2 Disagree 0 3 Neither 0 4 Agree	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. 	B) How would the organization like you to answer this item? 0 1 Strongly Disagree 0 2 Disagree 0 3 Neither 0 4 Agree 0 5 Strongly Agree	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure O ⁵ Very unsure
 2. "I usually arrive early for appointm A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. O ⁶ None of the above. O ⁷ I don't know 	B) How would the organization like you to answer this item? 0 1 Strongly Disagree 0 2 Disagree 0 3 Neither 0 4 Agree 0 5 Strongly Agree	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure O ⁵ Very unsure

3. "Sometimes I enjoy breaking the r	ules."	
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
O ² How confident and friendly a person is.	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.		
4. "I get annoyed when people chang	e things that work perfectly w	ell."
 4. "I get annoyed when people chang A) Which category best describes this item? 	e things that work perfectly we B) How would the organization like you to answer this item?	ell." C) How sure are you that you know how the organization would like you to answer this item?
 4. "I get annoyed when people changed when people changed	e things that work perfectly we B) How would the organization like you to answer this item? O ¹ Strongly Disagree	C) How sure are you that you know how the organization would like you to answer this item?
 4. "I get annoyed when people changed when people changed	e things that work perfectly we B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree	C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure
 4. "I get annoyed when people changed when people changed when people changed when people changed with the set of the set o	e things that work perfectly we B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither	 c) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure
 4. "I get annoyed when people changed with the person is the person is the person is the person is. 0 ¹ How adaptable a person is to different situations. 0 ² How confident and friendly a person is. 0 ³ How productive and focused on quality a person is. 0 ⁴ How a person reacts to being supervised. 	 e things that work perfectly we B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither O ⁴ Agree 	c) How sure are you that you know how the organization would like you to answer this item? O 1 Very sure O 2 Somewhat sure O 3 Neither sure nor unsure O 4 Somewhat unsure
 4. "I get annoyed when people changed with the person person	 e things that work perfectly we B) How would the organization like you to answer this item? 0 ¹ Strongly Disagree 0 ² Disagree 0 ³ Neither 0 ⁴ Agree 0 ⁵ Strongly Agree 	c) How sure are you that you know how the organization would like you to answer this item? O 1 Very sure O 2 Somewhat sure O 3 Neither sure nor unsure O 4 Somewhat unsure O 5 Very unsure
 4. "I get annoyed when people changed with the person person	 e things that work perfectly we B) How would the organization like you to answer this item? 0 ¹ Strongly Disagree 0 ² Disagree 0 ³ Neither 0 ⁴ Agree 0 ⁵ Strongly Agree 	c) How sure are you that you know how the organization would like you to answer this item? O 1 Very sure O 2 Somewhat sure O 3 Neither sure nor unsure O 4 Somewhat unsure O 5 Very unsure

5. "I catch on to things quickly."		
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ^I Very sure
 O ² How confident and friendly a person is. O ³ How productive and focused 	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.	[
6. "I do not have a good imagination.	››	
 6. "I do not have a good imagination. A) Which category best describes this item? 	" B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. 	" B) How would the organization like you to answer this item? O ¹ Strongly Disagree	C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused 	" B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither O ⁴ Agree 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. 	 B) How would the organization like you to answer this item? 0 ¹ Strongly Disagree 0 ² Disagree 0 ³ Neither 0 ⁴ Agree 0 ⁵ Strongly Agree 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure O ⁵ Very unsure
 6. "I do not have a good imagination. A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. O ⁶ None of the above. 	 B) How would the organization like you to answer this item? 0 ¹ Strongly Disagree 0 ² Disagree 0 ³ Neither 0 ⁴ Agree 0 ⁵ Strongly Agree 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure O ⁵ Very unsure

7. "I like working on several things a	t a time."	
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
 O² How confident and friendly a person is. O³ How productive and focused 	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.	l	
8. "Frequent interruptions and change	es in priority bother me."	
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
O ² How confident and friendly a person is. O ³ How productive and focused	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		

9. "I am always prepared."		
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
O ² How confident and friendly a person is. O 3 How productive and focused	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.		
10. "I tend to resist when people tell	me what to do."	••••••••••••••••••••••••••••••••••••••
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ^I Very sure
O ² How confident and friendly a person is. O ³ How productive and focused	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
$O = {}^{6}$ None of the above.		

11. I have little to say.	<u> </u>	
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
O ² How confident and friendly a person is. O ³ How productive and focused	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.		
12. "I avoid reading difficult material	99 10	
		C) How sure are you that you
A) Which category best describes this item?	B) How would the organization like you to answer this item?	know how the organization would like you to answer this item?
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree 	know how the organization would like you to answer this item? O ¹ Very sure
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree 	know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither 	know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither O ⁴ Agree 	know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure O ⁴ Somewhat unsure
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither O ⁴ Agree O ⁵ Strongly Agree 	know how the organization would like you to answer this item? 0 1 Very sure 0 2 Somewhat sure 0 3 Neither sure nor unsure 0 4 Somewhat unsure 0 5 Very unsure
 A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. O ⁶ None of the above. 	 B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither O ⁴ Agree O ⁵ Strongly Agree 	know how the organization would like you to answer this item? 0 1 Very sure 0 2 Somewhat sure 0 3 Neither sure nor unsure 0 4 Somewhat unsure 0 5 Very unsure

13. "A lot of supervisors just enjoy c	ontrolling people."	
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure
 O² How confident and friendly a person is. O³ How productive and focused 	O ² Disagree	O ² Somewhat sure
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure
O ⁶ None of the above.		
O ⁷ I don't know.		
14. "I typically am not very interested	in joining group activities."	· · · · · · · · · · · · · · · · · · ·
14. "I typically am not very interestedA) Which category best describesthis item?	in joining group activities." B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. 	 in joining group activities." B) How would the organization like you to answer this item? O ¹ Strongly Disagree 	C) How sure are you that you know how the organization would like you to answer this item?
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused 	 d in joining group activities." B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. 	 d in joining group activities." B) How would the organization like you to answer this item? O ¹ Strongly Disagree O ² Disagree O ³ Neither 	 C) How sure are you that you know how the organization would like you to answer this item? O ¹ Very sure O ² Somewhat sure O ³ Neither sure nor unsure
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. 	a in joining group activities." B) How would the organization like you to answer this item? O 1 Strongly Disagree O 2 Disagree O 3 Neither O 4 Agree	C) How sure are you that you know how the organization would like you to answer this item? O 1 Very sure O 2 Somewhat sure O 3 Neither sure nor unsure O 4 Somewhat unsure
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. 	a in joining group activities." B) How would the organization like you to answer this item? O 1 Strongly Disagree O 2 Disagree O 3 Neither O 4 Agree O 5 Strongly Agree	C) How sure are you that you know how the organization would like you to answer this item?O1O1Very sureO2Somewhat sureO3Neither sure nor unsureO4Somewhat unsureO5Very unsure
 14. "I typically am not very interested A) Which category best describes this item? O ¹ How adaptable a person is to different situations. O ² How confident and friendly a person is. O ³ How productive and focused on quality a person is. O ⁴ How a person reacts to being supervised. O ⁵ How good a person is at problem solving. O ⁶ None of the above. 	a in joining group activities." B) How would the organization like you to answer this item? O 1 Strongly Disagree O 2 Disagree O 3 Neither O 4 Agree O 5 Strongly Agree	C) How sure are you that you know how the organization would like you to answer this item?O1 Very sureO2 Somewhat sureO3 Neither sure nor unsureO4 Somewhat unsureO5 Very unsure

15. "I tend to ignore my duties."					
A) Which category best describes this item?	B) How would the organization like you to answer this item?	C) How sure are you that you know how the organization would like you to answer this item?			
O ¹ How adaptable a person is to different situations.	O ¹ Strongly Disagree	O ¹ Very sure			
 O² How confident and friendly a person is. O³ How productive and focused 	O ² Disagree	O ² Somewhat sure			
on quality a person is.	O ³ Neither	O ³ Neither sure nor unsure			
O ⁴ How a person reacts to being supervised.	O ⁴ Agree	O ⁴ Somewhat unsure			
O ⁵ How good a person is at problem solving.	O ⁵ Strongly Agree	O ⁵ Very unsure			
O ⁶ None of the above.					
O ⁷ I don't know.					

APPENDIX D

Unlikely Virtues

- 1. I always admit it when I make a mistake.
- 2. I never give up hope.
- 3. I know that anyone who tries can get a job.
- 4. I always know why I do things.
- 5. I never give up.
- 6. I Know immediately what to do.
- 7. I believe there is never an excuse for lying.
- 8. I always know what I am doing.
- 9. I am always ready to start afresh.
- 10. I have never engaged in gossip.
- 11. I will do anything for others.
- 12. I am always prepared.
- 13. I don't always practice what I preach.*
- 14. I have some bad habits.*
- 15. I have sometimes had to tell a lie.*
- 16. I am not always honest with myself.*
- 17. I am not always what I appear to be.*
- * denotes reverse keyed items

APPENDIX E Informed Consent

Project Title:	Selection Test Responding
Investigator's Names:	Anthony Boyce and Dr. Ann Marie Ryan
Description and Explanation of Procedures:	This study is investigating how people respond to pre-employment tests like the one you took before being hired. You will be asked to respond to a test similar to the one you took as an applicant to [organization name]. You will also be asked to answer some additional survey questions.
Benefits:	This study will last about one and a half hours (1.5 hours). You will receive your normal hourly wage for participation.
Risks and Discomforts:	None

This study is investigating how people respond to pre-employment tests like the one you took before being hired. Thank you for participating in this study! Participation is completely voluntary, and all of your answers will be completely CONFIDENTIAL. No one at [organization name] will ever see or have access to your individual responses to any of the questions in this study. Your privacy will be protected to the maximum extent allowable by law. All of the study materials will be taken off-site by Anthony Boyce at the end of each day so that no one at [organization name] will have access to them.

Participation is voluntary, you may choose not to participate at all, or you may refuse to participate in certain procedures or answer certain questions or discontinue your participation at any time without penalty or loss of benefits.

If you have any questions about this study, please contact Anthony Boyce (Michigan State University, Baker Hall Room 20, East Lansing, MI 48824; boyceant@msu.edu). If you have questions or concerns regarding your rights as a study participant or if you are dissatisfied with your treatment during this study you may contact – anonymously if you wish – Peter Vasilenko, Ph.D., Chair of the Michigan State University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, email: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824. A copy of this consent form will be available for you to take home.

Please remember that your answers during this study are completely **confidential** and absolutely no one at [organization name] will know or see your individual responses. Only the investigators will have access to your individual responses.

Your signature below indicates your voluntary agreement to participate in this study.

First and Last Name (please print)	Signature
Today's Date	

APPENDIX F <u>Protocol</u>

Thank you for agreeing to participate in this study. My name is Anthony Boyce and I'm a student at Michigan State University. The research project that you are going to participate in today is part of the requirements for me to graduate and has no con nection with [organization name] other than that they have allowed me to ask you to participate. This study is investigating how people respond to pre-employment tests like the one you took before being hired. You will be asked to respond to a test similar to the one you took as an applicant to [organization name]. You will also be asked to answer some additional survey questions. Note, that you are not in any way obligated to participate in this study. If you would prefer not to participate, you are welcome to leave now or at any time throughout the study.

[Begin handing out the first consent form.]

The consent form I am handing out states that you are agreeing to participate in this study voluntarily. Additionally, it states that all your answers today will be completely confidential. No one at [organization name] will ever have access to your responses in today's study. In order to ensure this, I will take all of your surveys with me when I leave at the end of the day. The consent form also informs you that the study takes approximately one and a half hours and that you will be paid your normal hourly wage for participation today. If you voluntarily agree to participate in this study today, please print your name, sign the form, and write in today's date in the appropriate boxes. Are there any questions?

[Wait until everyone has finished reading and signing the forms and collect them.] [Begin handing out the first questionnaire.]

109

I am now handing out a questionnaire that is similar to the one that you filled out as an applicant to [organization name]. In order to examine how people respond to these types of questionnaires, I need to be able to link your responses to this survey to the survey we will take in the next phase of the study, so please write your last and first name in the box provided.

Please answer all of the questions as honestly and accurately as you can. Remember, your responses to this survey are completely confidential. No one but me will ever see or know your individual answers and your answers will never be communicated to anyone at [organization name] for any reason. Your answers will be used for research purposes only and will have no effect on you or your employment at [organization name] in any way. Therefore, please answer the following questions as honestly as possible. When you are finished please sit quietly until everyone is finished.

[When everyone is finished, collect the tests.]

[Begin handing out the second questionnaire.]

I am handing out the next survey now. I need to be able to connect your responses to the last questionnaire to your responses on this one, so please put your first and last name in the box provided.

Again, please answer all of the questions as honestly and accurately as you can. Remember, your responses to this survey are completely confidential. No one but me will ever see or know your individual answers and your answers will never be communicated to anyone at [organization name] for any reason. Your answers will be used for research purposes only and will have no effect on you or your employment at [organization name] in any way. Therefore, please answer the following questions as honestly as possible.

In the questions on this survey, the term *selection test* refers to any questionnaire like the one you just took that is given to people when they apply for a job. The phrase *responding in a desirable manner* refers to slightly exaggerating your responses to a selection test in order to make yourself look good, responding to a question in a desirable way even if you do not think that it is completely true, or outright lying on your answers. You may begin. When you are finished please sit quietly until everyone is finished.

[When everyone is finished, collect the surveys.]

[Begin handing out the second consent form]

We have one more thing to do before we're done today. I am currently handing out a consent form similar to the one that you filled out at the beginning of the session. In addition to investigating how people respond to selection tests, I am also investigating how peoples' responses to these tests change over time and how these changes relate to job performance. In order to investigate these issues, it is necessary for me to have access to your applicant test responses as well as to have access to your job performance data. Once this information is linked to the surveys you took today, all names will be removed and any identifying information will be destroyed. If you agree to allow me to collect this information, please print your name, sign the form, and write in today's date. Also, please remember that no one at [organization name] will ever have access to any of the individual information that you provided today.

When I am done running these sessions, I will make sure that everyone who has participated is given a debriefing form that describes the purpose of the study in greater

111

detail. Please do not tell other employees about the questionnaires that you filled out today or anything else about this study. It is very important that when people participate they all have exactly the same information at the same steps in the study, so again please do not tell other employees about the details of this study. Please return the consent forms to me when you have completed filling them out. Thank you for participating in this study.

APPENDIX G

Informed Consent 2

Project Title:

Selection Test Responding

Investigator's Names:

Anthony Boyce and Dr. Ann Marie Ryan

In addition to understanding how people respond to selection tests, this study is also investigating how these responses change over time and how these changes relate to job performance. In order to investigate these issues, it is necessary for the investigators to have access to your applicant test responses as well as to have access to your job performance data.

I am asking for your authorization to release your applicant test scores and performance ratings to me for purposes of this investigation only. You may decline this request to have your scores released. The data released will be treated as confidential and no one but the investigators will have access to this information. Once this information is linked to the surveys you took today, **names will be removed and any identifying information destroyed**. No one at [organization name] will have access to the individual information you provided today. Your privacy will be protected to the maximum extent allowable by law.

If you have any questions about this study, please contact Anthony Boyce (Michigan State University, Baker Hall Room 20, East Lansing, MI 48824; boyceant@msu.edu). If you have questions or concerns regarding your rights as a study participant or if you are dissatisfied with your treatment during this study you may contact – anonymously if you wish – Peter Vasilenko, Ph.D., Chair of the Michigan State University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, email: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824. A copy of this consent form will be available for you to take home.

Your signature below indicates your voluntary agreement to allow the investigators to obtain both your applicant test responses and job performance ratings from your supervisor.

First and Last Name (please print)	Signature
Today's Date	

APPENDIX H

Debriefing Form

Selection tests like the one you took as an applicant are good at predicting job performance. Research has shown that people with certain characteristics are better at some types of jobs than others. By using tests like these, companies are attempting to make sure that the people they hire will like the job they are hired to do and will be good at it.

However, sometimes people respond differently when taking this type of test as an applicant than they do when they take this type of test for other purposes (for example, for research purposes) and some people respond differently in the same situation at different times (for example, six months later). When people respond differently in differently in different situations or at different times it is more difficult for these types of tests to accurately predict job performance.

This study is attempting to figure out why some people respond differently in different situations and at different times. Additionally, this study is attempting to figure out whether these changes in answers relate to job performance. By addressing these issues the study is attempting to improve the quality of these tests so that they will be more beneficial to both workers and companies.

If you would like to read more about how these types of questionnaires relate to job performance you can go to the library for the articles listed below. If you have any questions regarding this study please contact Anthony Boyce by email: <u>boyceant@msu.edu</u> or by mail: Michigan State University, Baker Hall Room 20, East Lansing, MI 48824.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. <u>Personnel Psychology</u>, <u>44</u>(1), 1-26.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. <u>Personnel Psychology</u>, <u>48</u>(4), 775-802.

APPENDIX I

Informed Consent

Project Title:	Selection Test Responding
Investigator's Names:	Anthony Boyce and Dr. Ann Marie Ryan
Description and Explanation of Procedures:	This study is investigating how people respond to pre-employment tests like the one administered to applicants at [organization name]. You will be asked to provide performance appraisal ratings for a number of employees you supervise.
Benefits:	The amount of time this task will require depends on how many participating employees you supervise, but should take no longer than 30 minutes.
Risks and Discomforts:	None

This study is investigating how people respond to pre-employment tests like the one administered to applicants at [organization name] and how responses change over time. Thank you for participating in this study! Participation is completely voluntary, and all of your performance ratings will be completely CONFIDENTIAL. No one at [organization name] will ever see or have access to the ratings you provide. Your privacy will be protected to the maximum extent allowable by law. All of the rating forms will be taken off-site by Anthony Boyce at the end of each day so that no one at [organization name] will have access to them.

Participation is voluntary, you may choose not to participate at all, or you may refuse to participate in certain procedures or answer certain questions or discontinue your participation at any time without penalty or loss of benefits.

If you have any questions about this study, please contact Anthony Boyce (Michigan State University, Baker Hall Room 20, East Lansing, MI 48824; boyceant@msu.edu). If you have questions or concerns regarding your rights as a study participant or if you are dissatisfied with your treatment during this study you may contact – anonymously if you wish – Peter Vasilenko, Ph.D., Chair of the Michigan State University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, email: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824. A copy of this consent form will be available for you to take home.

Please remember that the performance ratings you provide during this study are completely confidential and absolutely no one at [organization name] will know or see the performance ratings you provide. Only the investigators will have access to the performance ratings you provide.

Your signature below indicates your voluntary agreement to participate in this study.

First and Last Name (please print)	Signature
Today's Date	

REFERENCES

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), <u>Action control: From cognition to behavior</u>. Heidelberg: Springer.

Ajzen, I., & Fishbein, M. (1980). <u>Understanding attitudes and predicting social</u> <u>behavior</u>. Englewood Cliffs, NJ.: Prentice-Hall.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. <u>Journal of Experimental Social Psychology</u>, <u>22</u>(5), 453-474.

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. <u>Psychological Science</u>, 7, 32 - 39.

Arthur, W., Jr., Woehr, D. J., & Graziano, W. G. (2001). Personality testing in employment settings: Problems and issues in the application of typical selection practices. <u>Personnel Review, 30(6), 657-676</u>.

Bandura, A. (1997). <u>Self-efficacy: The exercise of control</u>. New York, NY, US: W. H. Freeman/Times Books/ Henry Holt & Co.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. <u>Personnel Psychology</u>, 44(1), 1-26.

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. <u>Journal of Applied Psychology</u>, 78(1), 111-118.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. <u>Journal of Applied</u> <u>Psychology, 81(3), 261-272</u>.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? <u>International Journal of Selection & Assessment, 9(1-2)</u>, 9-30.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality & Social Psychology, 51(6), 1173-1182.

Beck, L., & Ajzen, I. (1991). Predicting dishonest actions using the theory of planned behavior. Journal of Research in Personality, 25(3), 285-301.

Bedeian, A. G., & Day, D. V. (1994). Difference scores: Rationale, formulation, and interpretation. Journal of Management, 20, 695-698.

Birkeland, S., Manson, T., Kisamore, J., Brannick, M., & Liu, Y. (2003). <u>A meta-analysis of the difference between job applicants and non-applicants on personality measures</u>. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychologists, Orlando, FL.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. <u>Personnel Psychology</u>, 52(3), 561-589.

Christiansen, N. D., Goffin, R.D., Johnston, N.G., & Rothstein, M.G. (1994). Correcting the 16PF for faking: effects on criterion-related validity and individual hiring decisions. <u>Personnel Psychology</u>, 47, 847 - 860.

Christie, R., & Geis, F. L. (1970). <u>Studies in Machiavellianism</u>. New York: Academic Press Inc.

Cialdini, R. B. (in press). Managing social norms for persuasive impact. <u>Social</u> <u>Influence.</u>

Cohen, J. (1988). <u>Statistical power analysis for the behavioral sciences</u> (2nd ed.). New Jersey: Lawrence Erlbaum.

Collins, J. M., & Gleaves, D. H. (1998). Race, job applicants, and the Five-Factor Model of Personality: Implications for Black psychology, industrial/organizational psychology, and the Five-Factor Theory. <u>Journal of Applied Psychology</u>, 83(4), 531-544.

Costa, P. T., & McCrae, R. R. (1985). The NEO Personality Inventory Manual. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 24(1960), 349-354.

Donovan, J. J., Dwight, S.A., & Hurtz, G.M. (2003). An assessment of the prevalence, severity, and verifiability of entry-Level applicant faking using the randomized response technique. <u>Human Performance, 16(1), 81 - 106</u>.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). The validity of noncognitive measures decays when applicants fake, <u>Proceedings of the Academy of</u> <u>Management</u> (pp. 127 - 131). Cincinnati, OH. Drasgow, F., & Kang, T. (1984). Statistical power of differential validity and differential prediction analyses for detecting measurement nonequivalence. Journal of Applied Psychology, 69(3), 498-508.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. <u>Personnel</u> <u>Psychology</u>, 15(2), 13-24.

Dwight, S. A., & Donovan, J.J. (2003). Do warnings not to fake reduce faking? <u>Human Performance</u>, 16(1), 1 - 23.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt (Eds.), <u>Measuring and Analyzing Behavior in Organizations: Advances in Measurement and</u> <u>Data Analysis</u>. San Francisco: Jossey-Bass.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. Journal of Applied Psychology, 84(2), 155-166.

Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. <u>Journal of Applied Psychology</u>, <u>86(1)</u>, 122-133.

Etzioni, A. (1988). <u>The moral dimension: Toward a new economics</u>. New York, NY, US: Free Press.

Ford, J. K., MacCallum, R. C., Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. <u>Personnel</u> <u>Psychology</u>, <u>39</u>(2), 291-314.

Fox, J. A., & Tracy, P. E. (1986). <u>Randomized response: A method for sensitive</u> surveys. Beverly Hills, CA: Sage.

Frei, R. L., Snell, A. F., McDaniel, M. A., & Griffith, R. L. (1998). Using a within subjects design to identify the differences between social desirability and ability to fake. Paper presented at the 13th Annual Conference of the Society of Industrial and Organizational Psychologists, Dallas, TX.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. International Journal of Selection and Assessment, 11, 340-344.

Gordon, R. A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. Journal of Personality & Social Psychology, 71(1), 54-70.

Graham, M. A., Monday, J., O'Brien, K., & Steffen, S. (1994). Cheating at small colleges: An examination of student and faculty attitudes and behaviors. <u>Journal of</u> <u>College Student Development, 35(4)</u>, 255-260.

Haaland, D., & Christiansen, N. D. (1998). <u>Departures from linearity in the</u> relationship between applicant personality test scores and performance as evidence of response distortion. Paper presented at the 22nd Annual International Personnel Management Association Assessment Council Conference, Chicago, IL.

Harrison, D. A. (1995). Volunteer motivation and attendance decisions: Competitive theory testing in multiple samples from a homeless shelter. Journal of Applied Psychology, 80(3), 371-385.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. <u>Personnel Psychology</u>, 57, 639-683.

Hogan, R., & Hogan, J. (1992). <u>Hogan personality inventory manual</u>. Tulsa: Hogan Assessment Systems.

Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. <u>American Psychologist, 43(8), 621-626</u>.

Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), <u>Handbook of industrial and organizational psychology</u> (2 ed., Vol. 2, pp. 873-919). Palo Alto, CA: Consulting Psychologists Press, Inc.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. <u>Human Performance, 11(2)</u>, 209 - 244.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & et al. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75(5), 581-595.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. <u>Structural Equation</u> <u>Modeling, 6(1), 1-55</u>.

Hurd, J. M., Barrett, G. V., Miguel, R. F., Tan, J. A., & Lueke, S. B. (2001). When do response distortion scales reflect faking? A meta-analysis. Paper presented at the 16th Anual Conference of the Society for Industrial and Organizational Psychologists, San Diego, CA.

International Personality Item Pool. (2001). A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences. (http://ipip.ori.org/): Internet Web Site.

Kanji, G. K. (1993). 100 Statistical Tests. London: SAGE Publications.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? Journal of Personality & Social Psychology, 70(5), 1037-1051.

Komar, S., Theakston, J., Brown, D. J., & Robie, C. (2005). Faking and the validity of personality: A monte carlo investigation. <u>Manuscript submitted for publication.</u>

Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. Journal of Consulting & Clinical Psychology, 43(1), 48-55.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. <u>Psychological Bulletin, 107(1)</u>, 34-47.

Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies, and bullshit: Understanding faking and its effects on selection decisions using theory, field research, and simulation. In J. M. Brett & F. Drasgow (Eds.), <u>The Psychology of Work:</u> <u>Theoretically Based Empircal Research</u> (pp. 253 - 276). Mahwah, NJ.: Lawrence Erlbaum Associates.

Lueke, S. B., Snell, A. F., Elllingworth, A. J., & Paidas, S. M. (2001). <u>An</u> <u>empirical test of an interactional model of faking</u>. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychologists, San Diego, CA.

McFarland, L. A. (2000). <u>Toward an integrated model of applicant faking</u>. Unpublished dissertation. Michigan State University, East Lansing.

McFarland, L. A. (2002). <u>Consequences of warning against faking on a</u> <u>personality test.</u> Paper presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychologists, Toronto, ON.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. Journal of Applied Psychology, 85(5), 812-821.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. <u>Psychometrika</u>, 58(4), 525-543.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., III. (2003a). Faking and selection: Considering the use of personality from select-in and select-out perspectives. Journal of Applied Psychology, 88(2), 348-355.

Mueller-Hanson, R. A., Heggestad, E. D., & Thornton, G. C. (2003b). <u>Individual</u> <u>differences in impression management: An exploration of the psychological processes</u> <u>underlying faking</u>. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychologists, Orlando, FL.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. <u>Human</u> <u>Performance, 11(2-3), 245-269</u>.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. <u>Journal of Applied</u> <u>Psychology, 81(6), 660-679</u>.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive metaanalysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. <u>Journal of Applied Psychology</u>, 78(4), 679-703.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. <u>Personnel Psychology</u> <u>Special Issue: Theory and literature, 48</u>(4), 775-802.

Pandey, J., & Rastogi, R. (1979). Machiavellianism and ingratiation. Journal of Social Psychology, 108(2), 221-225.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. Journal of Personality & Social Psychology, 46(3), 598-609.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleiner & J. S. Wiggins (Eds.), <u>Personality assessment via</u> <u>questionnaire</u> (pp. 142 - 165). New York: Springer.

Reynolds, D. H., Sinar, E. F., & Haaland, D. E. (2003). <u>Non-cognitive testing in</u> <u>practice: Effects of preparation on score characteristics and subgroup differences.</u> Paper presented at the 18th Annual Conference of the Society of Industrial and Organizational Psychologists, Orlando, FL.

Robie, C., Zickar, M. J., & Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. <u>Human</u> <u>Performance, 14(2), 187-207</u>.

Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. <u>Multivariate Behavioral Research</u>, 13(4), 403-418.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. <u>Psychological Bulletin, 92</u>, 726-748.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. <u>Journal of Applied Psychology</u>, 83(4), 634-644.

Schifter, D. E., & Ajzen, I. (1985). Intention, perceived control, and weight loss: An application of the theory of planned behavior. <u>Journal of Personality & Social</u> <u>Psychology</u>, 49(3), 843-851.

Schlenker, B. R. (1980). Impression Management. Monterey, CA: Brooks/Cole.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. <u>Journal of Applied Psychology</u>, <u>78</u>(6), 966-974.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), <u>International Review of Industrial and Organizational Psychology</u> (pp. 115 - 140). New York: Wiley.

Schmitt, N., & Oswald, F. L., (in press). The impact of corrections for faking on the validity of noncognitive measures in selection settings. <u>Journal of Applied</u> <u>Psychology.</u>

Schoorman, F. D., Bobko, P. & Rentsch, J. (1991). The role of theory in testing hypothesized interactions: An example from the research on escalation of commitment. Journal of Applied Social Psychology, 21, 1338-1355.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. Journal of Applied Psychology, 87(2), 211-219.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. Journal of Applied Psychology, 86(2), 304-315.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: integrating studies of deception. <u>Human Resource Management Review</u>, 9(2), 219 - 242.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. <u>Psychological Bulletin, 124(2), 240-261</u>.

Stark, S., Chernyshenko, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. Journal of Applied Psychology, 86(5), 943-953.

Tellegen, A. (in press). <u>MPQ (Multidimensional Personality Questionnaire):</u> <u>Manual for administration, scoring, and interpretation</u>. Minneapolis: University of Minnesota Press.

Tisak, J. & Smith, C. S. (1994). Defending and extending difference score methods. Journal of Management, 20(3), 675-682.

Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. <u>Personality & Individual Differences</u>, 23(1), 117-124.

Triandis, H. C. (1977). Interpersonal behavior. Monterey, CA.: Brooks/Cole.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. <u>Educational & Psychological Measurement</u>, <u>59</u>(2), 197-210.

Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? <u>International Journal of Selection & Assessment</u>, 9(4), 277-289.

Wayne, S. J., & Liden, R. C. (1995). Effects of impression management on performance ratings: A longitudinal study. <u>Academy of Management Journal, 38(1)</u>, 232-260.

Weekley, J. A., Ployhart, R. E., & Harold, C. (2003). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychologists, Orlando, FL.

Weichmann, D. (2000). <u>Applicant reactions to novel selection tools</u>. Unpublished thesis. Michigan State University, East Lansing.

Weiner, J. A., & Gibson, W. M. (2000). <u>Practical effects of faking on job</u> <u>applicant attitude test scores.</u> Paper presented at the 15th Annual Conference of the Society of Industrial and Organizational Psychologists, New Orleans, LA.

Whyte, W. H. (1957). The organization man. Garden City, NY: Doubleday.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. <u>Academy of Management Review</u>, 12(2), 250-264.

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. <u>Applied Psychological Measurement, 20(1)</u>, 71-87.

