2009


Esteban Tristan
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the Industrial and Organizational Psychology Commons

Repository Citation
https://corescholar.libraries.wright.edu/etd_all/915

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
MEASURING APPLICANT FAKING WITH JOB DESIRABILITY:
PREVALENCE, SELECTION, AND MEASUREMENT
ISSUES IN AN APPLIED SAMPLE

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

By

ESTEBAN TRISTAN
B.A., University of Rhode Island, 1999
M.S., Wright State University, 2003

2009
Wright State University
I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Esteban Tristan ENTITLED BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Corey E. Miller, Ph.D.
Dissertation Director

Dan Weber, Ph.D.
Director, HF-IO Psychology
Ph.D. Program

Joseph F. Thomas, Jr., Ph.D
Dean, School of Graduate Studies

Committee on Final Examination

Martin Gooden, Ph.D.

David LaHuis, Ph.D.

Debra Steele-Johnson, Ph.D.
ABSTRACT


Research has found that the use of social desirability scales to measure faking is problematic. The current study employed a job desirability scale consisting of job-specific bogus biographical items as an alternative faking measure in an applied setting. Using a 2 (applicants versus incumbents) x 2 (sales versus managers) design, participants (N = 958), participants completed a set of personality, social desirability, and job desirability measures. Results indicated that applicants outscored incumbents on personality measures. However, the effect size for conscientiousness was larger for the manager job whereas the effect size for extraversion was larger for the sales job, indicating a job-specific pattern of faking. Furthermore, applicants faked most on bogus items that were specific to the job they were applying for (sales vs. manager). Applicants who faked on the job desirability scale also systematically increased their chances of being hired over non-fakers. Job desirability scores displayed weak to moderate correlations with personality, social desirability, and job experience, although these varied somewhat by job type. The results have various implications for: non-cognitive test usage in selection settings, assumptions regarding incumbent test scores, the unique challenge of faking by sales applicants, and the further understanding of job-specific faking.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>Faking is Possible</td>
<td>6</td>
</tr>
<tr>
<td>Can Faking be Reduced or Eliminated? Previous Findings</td>
<td>7</td>
</tr>
<tr>
<td>Faking is Prevalent</td>
<td>12</td>
</tr>
<tr>
<td>Effects of Faking on Construct Validity</td>
<td>20</td>
</tr>
<tr>
<td>Effects of Faking on Criterion Validity – Faking Does Not Lower Validity</td>
<td>25</td>
</tr>
<tr>
<td>Effects of Faking on Criterion Validity – Faking Does Lower Validity</td>
<td>27</td>
</tr>
<tr>
<td>Problems with Validity Coefficients as Indicators of Faking Effects</td>
<td>31</td>
</tr>
<tr>
<td>Fakers “Rise to the Top” Regardless of Stable Validities</td>
<td>34</td>
</tr>
<tr>
<td>Faking is Not Equivalent to Social Desirability – Theory</td>
<td>39</td>
</tr>
<tr>
<td>Faking is Not Equivalent to Social Desirability – Research Findings</td>
<td>42</td>
</tr>
<tr>
<td>A Better Description of Applicant Faking – Job Desirability</td>
<td>46</td>
</tr>
<tr>
<td>The Measurement of Job Desirability</td>
<td>48</td>
</tr>
<tr>
<td>Job Desirability Measures – More Sensitive to Faking than Social Desirability</td>
<td>51</td>
</tr>
<tr>
<td>Conclusion</td>
<td>56</td>
</tr>
<tr>
<td>The Proposed Research – Purpose and Hypotheses</td>
<td>56</td>
</tr>
</tbody>
</table>
# Methodist University - Library

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>III. Method</td>
<td>67</td>
</tr>
<tr>
<td>Participants</td>
<td>67</td>
</tr>
<tr>
<td>Measures</td>
<td>68</td>
</tr>
<tr>
<td>Procedure</td>
<td>71</td>
</tr>
<tr>
<td>I. RESULTS</td>
<td>74</td>
</tr>
<tr>
<td>Demographics</td>
<td>74</td>
</tr>
<tr>
<td>Scale Reliabilities</td>
<td>74</td>
</tr>
<tr>
<td>Tests of Hypotheses</td>
<td>75</td>
</tr>
<tr>
<td>II. DISCUSSION</td>
<td>86</td>
</tr>
<tr>
<td>Limitations of the Current Study</td>
<td>108</td>
</tr>
<tr>
<td>Summary</td>
<td>115</td>
</tr>
<tr>
<td>III. REFERENCES</td>
<td>139</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table

1a. Demographic Frequencies for Study Sample Only …………………… 118
1b. Demographic Frequencies for All Participants …………………… 119
2a. Means, Standard Deviations, and Correlation Matrix of Study Variables – Full Sample …………………………………………………… 120
2b. Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Applicants and Incumbents ………………………………. 121
2c. Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Applicants and Incumbents ………………………….. 122
2d. Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Applicants …………………………………………………… 123
2e. Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Applicants …………………………………………………… 124
2f. Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Incumbents …………………………………………………… 125
2g. Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Incumbents …………………………………………………… 126
3a. Differences Between Applicant and Incumbent Personality Scores for Sales Job ……………………………………………………………….. 127
3b. Differences Between Applicant and Incumbent Personality Scores for Manager Job ……………………………………………………………….. 127
3c. Differences Between Applicant and Incumbent Scores on Job Desirability and Social Desirability for Sales Job…………………………………… 128
3d. Differences Between Applicant and Incumbent Scores on Job Desirability and Social Desirability for Manager Job…………………………………… 128
4a. Proportion of Sales Applicant Fakers Selected Under Varying Selection Ratios Based on Top-Down Ranking of Personality Scores……………… 129
4b. Proportion of Manager Applicant Fakers Selected Under Varying Selection Ratios Based on Top-Down Ranking of Personality Scores 129
5. Within-Subject Differences on Sales versus Manager Job Desirability Scores for Applicants………………………………………… 130
6. Between-Subject Differences on Sales versus Manager Job Desirability Scores for Applicants………………………………………… 131
7. Mixed-Design Analysis of Variance on Job Desirability Scores for Sales and Manager Scale Versions by Job Group…………………………………… 132
8a. Pearson’s r Correlations Between Job Desirability and Social Desirability Scale in Sales and Manager Applicant Samples…………………………………… 133
8b. Spearman’s Rho Correlations Between Job Desirability and Social Desirability Scale in Sales and Manager Applicant Samples…..…… 133
9a. One-Sample Tests on Sales Job Desirability Scores for Sales Incumbents and Applicants……………………………………………………… 134
9b. One-Sample Tests on Manager Job Desirability Scores for Manager Incumbents and Applicants……………………………………………………… 134
Table 10. Parametric and Non-Parametric Correlations Between Job Desirability and Job Experience in Sales and Manager Applicant Samples………………………………………135

Table 11a. Item Analysis of Sales Job Desirability Scale - Sales and Manager Applicant Samples…………………………………………………………………………….136

Table 11b. Item Analysis of Manager Job Desirability Scale - Sales and Manager Applicant Samples……………………………………………………………………137

Appendix
A. Goldberg’s (1999) IPIP Big-Five Factor Scales 156
B. Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991) 158
C. Job Desirability Scales 160
D. Assessment Instructions 162
LIST OF FIGURES

Figure

1. Interaction Effects of Job Group X Job Desirability Scale Version on Job Desirability Scores …………………………………………………………………………………….. 138
ACKNOWLEDGEMENTS

First and foremost I would like to thank God for all the blessings He has provided which have made this possible. I have many individuals to thank for their support and encouragement throughout the years, without whose contributions, this work would not have been possible. My dear wife Stephanie, you have stood by me since my first day as a graduate student and I simply could not have accomplished this without you. The sacrifices you have made for me time and time again have made this accomplishment possible. Mom and Dick, thank you for teaching me your work ethic and never letting me give up my dream of obtaining a doctorate. I hope to have as much faith in my children as you have had in me. Papi and Livia, I will always remember and teach my children this advice you gave me: “estudie y estudie, y cuando se cansa de estudiar…pues estudie más.” Thank you for always pushing me to aim higher. And to my grandparents, thank you, Moni and Dad for your ongoing support and for always stressing the value of education. I owe a world of thanks to my advisor and mentor, Dr. Corey Miller, who taught me so much not only I/O Psychology, but about being a professional. I am indebted to you for your guidance and patience and for making the time this past year to help me accomplish this endeavor whilst living on the other side of the world. And last but not least, I give many thanks to Mom and Dad Bruce, the best mother and father-in-law one could ever have. Your love, support, and willingness to lend a hand have helped me, and our family, in countless ways. Many, many thanks to all of you, and to all the family, friends, and co-workers, who have supported me in this endeavor.
I. INTRODUCTION

Various research findings in the last two decades support the claim that personality tests add predictive utility for personnel selection. Large scale meta-analytic reviews by Barrick, Mount, and Judge (2001) and Hurtz and Donovan (2000) have concluded that measures of conscientiousness and neuroticism were significantly correlated with job performance measures across various occupation types. They also found that other Big Five personality factors, such as extraversion and agreeableness significantly predicted performance in specific job settings, such as sales and team jobs. This confirmed findings from earlier meta-analyses which reported that personality tests were predictive of job performance for various job types (Barrick & Mount, 1991; Hough 1992; Salgado, 1997; Tett, Jackson, & Rothstein, 1991). Similar findings have been reported for other non-cognitive measures, such as integrity tests (Ones, Viswesvaran, & Schmidt, 1993), biodata (Hunter & Hunter, 1984; Reilly & Chao, 1982) and customer service measures (Frei & McDaniel, 1998).

Despite these advantages, much concern has been raised about the susceptibility of non-cognitive tests to faking, or response distortion. Numerous studies have established that such tests can be faked easily by respondents (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Viswesvaran & Ones, 1999; Zickar & Robie, 1999). Most of the studies documenting this susceptibility to faking, however, were laboratory studies employing designs which instructed respondents to fake. Although some research has suggested that
faking is not prevalent in actual applicant settings (Hogan, Barrett, & Hogan, 2007), evidence of intentional response distortion in applied selection situations does exist (Griffith, Chemielowski, and Yoshita, 2007; Anderson, Warner, & Spencer, 1984). This debate has been and continues to be a source of much conflict in the faking literature and it will only be resolved through the continued use of actual applicant samples in faking research.

These conflicting findings also have led to a debate over the effects of faking on the construct validity of personality tests. While some studies indicate that faking does not alter the factor structure of personality tests (Bradley & Hauenstein, 2004; Robie, Zickar, and Schmit, 2001) other research indicates that faking does detrimentally alter factor structure (Van Iddekinge, Raymark, Eidson, & Putka, 2003, Weekley, Ployhart, & Harold, 2003).

Ultimately, however, faking is not a concern to practitioners if it does not reduce the predictive, or criterion, validity of personality measures in predicting job performance. This issue has been an even larger source of debate. Ones, Viswesvaran and Reiss (1996) performed a meta-analysis which found that controlling for social desirability did not improve validity of personality measures overall. Other research also has concluded that corrections for social desirability have no effect on criterion validity (Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough 1998). In addition, some studies comparing applicant to incumbent samples have found very similar criterion validity estimates regardless of sample type, suggesting that although applicants may inflate their scores on personality measures, it does result in lower validity in predicting their job performance (Weekley et al., 2003).
On the opposing side of this dispute, however, there are researchers who report that faking does reduce criterion-related validity in selection (Harold, McFarland, & Weekley, 2003; Rosse, Stecher, Miller, & Levin, 1998). Due to the widespread use of personality testing in actual selection settings today, this matter has important implications. If personality tests cannot predict job performance when applicants fake on them, personality testing may result in a loss of valuable time and money.

These conflicting results may be due largely to the way in which faking is conceptualized by different researchers. Traditionally, faking has usually been operationally defined as social desirability (Hough et al., 1990; Barrick & Mount, 1996). This has led researchers to use measures of social desirability to identify respondents who fake on personality tests and then statistically correct personality scores for social desirability. However, social desirability is a response bias toward giving an overall good social impression (Furnham, 1986), which, is not entirely deceitful or conscious (Paulhus, 1984). Since intentional deception is precisely the type of faking that practitioners are most concerned about, however, social desirability scales may not be an appropriate measure of faking (Kluger & Colella, 1993; Pannone, 1984). Moreover, recent findings indicate that job applicants tend to fake in a manner which is complex and highly specific to the job they are applying for (Birkeland, Manson, Kissamore, & Brannick, 2006; Raymark, Shilobod, & Steffensmeier, 2004), which is inconsistent with the nature of social desirability (Furnham, 1986) thus prompting the search for new constructs to model this job-specific faking and appropriate measures to detect it (Carroll, Jones, & Sulsky, 2004; Miller & Tristan, 2002).
Clearly, additional research is necessary in order to resolve the issues surrounding:

1) the actual prevalence of faking in applied settings, 2) the extent to which faking attenuates construct and criterion validity, and 3) which conceptualizations and measures of faking will be most effective in identifying fakers. The current study seeks to address these main questions and related issues.
II. LITERATURE REVIEW

Faking has been given many different names in the literature. It has been referred to as response distortion, dissimulation, impression management, claiming unlikely virtues, lying, inflation bias, and self-enhancement (Anderson et al., 1984; Edwards, 1957; Hough et al., 1990; Hough & Paulin, 1994; Lautenschlager, 1994; Paulhus, 1984). Extensive research has been conducted on faking, yet the exact nature of faking and its effects on personality tests remains unclear. Thus far, some of the most pivotal questions that have been addressed in faking research are: 1) is faking possible? 2) Can faking be reduced or eliminated? 3) Is faking prevalent in applied settings? 4) Does faking lower the construct and criterion validity of personality measures? 5) Is faking equivalent to social desirability, and if it is not, 6) are there more effective methods of detecting fakers than social desirability measures?

Currently, however, three of these issues are at the center of debate in the field: the prevalence of faking in applied settings, whether faking affects criterion validity, and whether social desirability measures effectively measure faking. The current literature review will address all of these topics and argue that faking is prevalent in the applied world, it adversely affects criterion validity, and it is not adequately measured by social desirability. Instead, a more recent alternative measure of faking, job desirability, is proposed as a more effective measure of faking and is argued to be specifically designed to detect faking in actual selection applicant situations.
Faking is Possible

The ability of subjects to fake on personality tests has been documented consistently over years of research. Early studies, such as Meehl and Hathaway’s (1946) study on the Minnesota Multiphasic Personality Inventory (MMPI) K-Factor, have found that subjects can present themselves in a manner that significantly influences their scores. Subjects portrayed themselves in too favorable a light or tended to be overly honest and self-critical. These tendencies were described by the authors as deliberate and conscious efforts to fake. As Meehl and Hathaway (1946) noted, however, various earlier studies had warned about the possibility of faking and its detrimental effects on personality measures (Allport, 1928; Bernreuter, 1933; Kelly, Miles, & Terman, 1936; Vernon, 1934). The fakability of personality tests has become increasingly apparent, as studies by Borislow (1958), Dicken (1959), and Orpen (1971) showed that subjects could present themselves deliberately in a more favorable light for certain purposes on Edwards' (1959) Personal Preference Schedule (EPPS). Strengthening the findings was the fact that this test was designed specifically to reduce socially desirable responding (Hogan, 1991).

Additional studies have continued to examine the susceptibility of personality tests to faking. A study by Hinrichsen (1972) used college students as subjects and divided them into two groups; one was instructed to answer honestly whereas the other was told to fake in the direction of a successful top manager on the Gordon Personality Inventory (GPI). Results indicated that those told to fake scored significantly higher than those in the honest condition. A similar study instructed college undergraduates to either "answer honestly", "fake good", or "fake bad" on the FIRO-B and found the test could be faked when subjects were instructed to do so (Hinrichsen, Gryll, Bradley, & Katahn, 1975).
Dunnette, Koun, and Barber (1981) concluded that the Eysenck Personality Inventory also was fakable under instructions to fake good (make a good impression) or to fake bad (make a poor impression). Further evidence came from a study involving personnel in various military settings, which found that respondents successfully distorted their self-descriptions in both a favorable and unfavorable manner when instructed to do so (Hough et al., 1990).

Meta-analysis studies have provided evidence on a larger scale that respondents can distort their scores substantially. Viswesvaran and Ones (1999) provided meta-analytic findings which reported that respondents could change their scores by over 0.50 standard deviations on several Big Five personality inventories and integrity tests. Alliger & Dwight (2000) also reported that personality and integrity test scores could be increased by subjects upon instructions to fake good or upon being coached on how to fake. Subjects increased their integrity test scores by as much as 1.32 standard deviations and increased their personality test scores by as much as 0.38 standard deviations. In yet another meta-analysis, which examined data from 25 studies of the MMPI, researchers concluded that subjects with symptoms of psychopathology could underreport (fake) their scores by as much as one standard deviation on the average (Baer, Wetter, & Berry, 1992). Thus the evidence strongly suggests that people can easily fake personality tests when instructed to do so.

Can Faking Be Reduced or Eliminated? Previous Findings

Attempts to make non-cognitive measures less susceptible to faking have a long history in industrial and organizational psychology, however, this history is beyond the scope of this review. Nevertheless, it is possible to review some currently used methods
used to deter faking, along with some novel alternative approaches which seem promising thus far.

One method which has been accepted for some time as an effective means to deter faking is simply warning applicants not to fake (e.g., Hough, 1998). A review by Dwight and Donovan (2003) highlights that although some studies have shown that warnings do reduce faking, other studies suggest that warnings have negligible effects. The authors reviewed the extant research and identified 15 studies on faking and warnings, 10 of which had adequate information necessary for estimating effect sizes. They found a sample-weighted mean effect size of 0.23 standard deviations, which they classified as a “weak” effect.

Dwight and Donovan (2003) went on to discuss a possible reason for the inconsistency in results – information conveyed in the warning. Warnings tend to either convey that 1) those who fake can be identified, 2) what the negative consequences of faking on the test will be, or 3) both of these messages simultaneously. Upon analyzing the effect that this information factor had on the same studies, they concluded that it had a substantial effect; the average effect size for studies only warning about identification was nearly zero (0.01 standard deviations) whereas the average effect size for studies warning about the consequences of faking was 0.30. Combining these two information pieces resulted in an average effect size of 0.25. Although only 10 studies were included in the review, results indicate that information conveyed in a warning not to fake is important. A subsequent laboratory study by the authors tested the effect of these types of information in a warning not fake. Results indicated that all three types of warnings (identification only, consequences only, or both) reduced scores on a personality measure.
However, only the warning that included both identification and consequences information produced scores that were significantly lower than scores in the no-warning group. Harold, McFarland, Dudley, and Odin (2004) used a warning which conveyed both these pieces of information. It also resulted in significantly lower personality and social desirability scores for those who were warned than for those not warned. Thus there is evidence that warnings may effectively deter faking if they convey that faking can be detected and that this will result in negative consequences, such as an applicant being removed from the applicant pool.

Another proactive technique used to reduce applicant faking is the construction of ipsative measures of personality. Ipsative measures have existed for many years and were specifically designed to control socially desirable responding, or the tendency to give overly positive self-descriptions (this topic will be discussed in further detail in a later section). An ipsative measure presents respondents with response options that are equal in social desirability, thus their choice should not be influenced by social desirability. Respondents must choose only one option that is the “most true” of them and choose one which is the “least true” of them in their everyday behavior. A major underlying assumption is that when respondents are forced to choose among a group of equally positive options (usually four), the option that is actually most true of them will be perceived as the best choice in describing themselves (Bowen, Martin, and Hunt, 2002).

Research has found that, in fact, ipsative measures are more resistant to faking than normative measures (Bowen et al., 2002). However, major psychometric differences exist between ipsative and normative measures, which has led to a preference for normative measures in North America (Bowen et al., 2002). One of these differences is
that in ipsative formats, the sum of all scales comprising the total test adds to a constant, such that every respondent receives the same score on the total test. This limits any elevation of scores on the total measure; only the subscales can differ in total score. It is this psychometric requirement which makes a measure truly ipsative. Such a scoring system, in effect, limits the investigator to intra-individual differences only, whereby a person’s relative strengths and weaknesses on a given construct can be compared. It precludes any inter-individual comparisons, however. Normative and ipsative measures also differ in many other aspects, such as item structure, presentation, questionnaire instruction, answering, and statistical assumptions. Together, these differences have resulted in the infrequent use of ipsative measures in North American I/O psychology, although the same is not true of Asia and most of Europe (Bowen et al., 2002).

Other personality measurement techniques which capitalize upon a forced-choice format have been proposed. An example is James’ (1998; 1999) Conditional Reasoning Test (CRT), which was developed as an alternative to self-report measures of personality. Although it also features a forced-choice format, it differs from ipsative measures in many key elements, including its theory. The Conditional Reasoning approach measures personality indirectly by measuring how a person perceives a particular situation. It is assumed that people have different latent motives which lead them to view, or frame, the world differently and that certain situations make these latent motives highly salient. The way a person frames their world leads them to construct an idiosyncratic logical reasoning in order to explain their behavior. This reasoning can be considered “conditional” because it relies heavily upon a person’s latent motives. Therefore, the Conditional Reasoning system assesses a person’s latent motives (personality) by examining their
logical reasoning when they solve problems. The CRT thus appears to be a problem-solving test to the test taker, who does not realize that what is actually being measured is their personality (James, 1998; 1999; 2000).

It is precisely this design aspect which makes the CRT more resistant to faking. LeBreton, Burgess, and James (2000) found that participants under instructions to fake good were significantly less able to fake the CRT than a more transparent measure of personality. Hence, measuring personality in a more indirect fashion seems to reduce faking.

Using a similar model of indirect measurement, Leasher, Miller, April, Gildea, Rees, Schwartz, and Tristan (2004) developed a “scenario-based” test format in which participants were presented with a scenario which is intended to activate latent motives. Participants then select among certain alternatives in response to the scenarios. Each alternative represents a possible reaction to the scenario. When instructed to fake for a hypothetical job which they wanted, this scenario-based measure of conscientiousness, along with the IPIP Conscientiousness scale, the authors found that the scenario based measure was substantially less fakable than the IPIP. It appeared that participants were unable to decipher which responses to the situations would increase their scores and give them the best chance to be hired.

Although biodata items are qualitatively different from personality items, research indicates that they are nonetheless fakable as well (Lautenschlager, 1994). However, research indicates that the more verifiable a response to a biodata question is, the less likely it is to be faked (Becker & Colquitt, 1992; Donovan, Dwight, & Hurtz, 2003). For example, a strategy developed by Schmitt and Kunce (2002) called “item elaboration” has
shown some promise in deterring response distortion on biodata measures. Item elaboration requires respondents to elaborate on their answers to biodata questions. For example, if an item asked the respondent how many times he or she led class discussion during their senior year of high school, after endorsing one of the available frequency response options, the respondent would then be asked to list the classes and discussion topics that they actually led. This example was taken from a study by Schmitt, Oswald, Kim, Gillespie, Ramsay, and Yoo (2003), who replicated Schmitt and Kunce’s (2002) earlier finding that subjects score significantly lower on biodata measures when they are required to elaborate on their responses. Thus, the requirement of elaborating upon one’s experiences on biodata measures appears to deter faking on this type of non-cognitive measure. This is in agreement with other findings (Lautenschlager, 1994).

Faking is Prevalent

Although the evidence clearly shows that faking is possible, it is still strongly debated whether applicants in actual selection settings actually engage in faking to any significant extent. Unfortunately, findings regarding the actual prevalence of faking have been contradictory. Abrahams, Neumann, & Githens (1971) found no significant differences in test scores when comparing administrations of the Strong Vocational Interest Blank (SVIB) under laboratory versus actual selection conditions, indicating that in actual selection conditions, applicants did not fake the SVIB. A similar study by Orpen (1971) compared students' change of scores on the Edwards Personal Preference Schedule (EPPS) with those of actual applicants for a clerical position. Both students and selection applicants responded under two conditions - as either part of a selection procedure or anonymous research participation. Although students changed their scores significantly
from the first administration to the second, the job applicants' scores were similar in both conditions, indicating little tendency of applicants to fake on the EPPS in an actual selection setting.

Schwab and Packard (1973) found similar results in a study using the Guilford Personality Inventory (GPI) and Guilford Personality Profile (GPP). No significant mean score differences were found between a group who was told that results would be used as part of the selection process and a second group that was hired then asked to complete the tests for unrelated research purposes only. Such low prevalence of faking has been documented also by Becker and Colquitt (1992) who found that actual selection applicants faked biodata items to a much lesser degree than did incumbents instructed to fake. Applicant responses were more similar to those of incumbents told to answer honestly. Ryan and Sackett (1987) also reported no significant differences between honest and simulated applicant respondents on a pre-employment honesty measure. Hough et al.’s (1990) study involving military subjects found that the personality test scores of newly sworn-in military personnel who took the test as a means for making decisions about their future military careers were very similar to the scores of incumbents. Therefore, the authors concluded that faking on personality tests has a low prevalence rate. Thus, there is strong evidence indicating that faking is not prevalent in applied settings, and is therefore not a cause for concern.

Despite such strong evidence to the contrary, however, many studies have found that faking in actual selection settings is highly prevalent. Early research on faking has provided support for this argument. Green (1951) reported that applicants for police officer positions responded in a more favorable style than those who were already
selected. In a study on fakability of emotional stability measures, Heron (1956) observed that applicants for a bus conductor position faked a self-report inventory in a selection situation but did not fake when they were tested again in a post-selection situation for research purposes only. Bass (1957) investigated the faking behavior of sales applicants on a forced-choice personality instrument and found that applicants scored significantly higher on three of four personality sub-scales than did incumbents. Elliot (1976) concluded that bank and industrial managers from Britain and Ireland significantly faked the 16PF questionnaire during a selection situation.

Further evidence of faking in a job applicant setting was provided by Pannone (1984), who found that over one-third of applicants faked a questionnaire of job-relevant skills. Results similar to Pannone's (1984) were found in another experiment where nearly one-half of the applicants also faked on a job-relevant skills questionnaire (Anderson et al., 1984). Griffith, Chmielowski, Snell, Frei and McDaniel (2000) found 55% of applicants at a temporary employment agency had scored at least one-half a standard deviation higher on a measure of conscientiousness when they took the test in an actual selection situation compared to when they took the test under an “honest” condition anonymously one month later for research purposes only. More importantly, the authors reported significant rank-order changes in applicants’ test scores as a function of testing situation, which indicates that faking can increase an applicant’s chances of being hired. The literature has documented additional evidence of response distortion in applied settings (Barrick & Mount, 1996; Costello, Schneider, and Schoenfeld, 1993; Elliot, 1981; Hough, 1995; 1996).
Rosse et al. (1998) conducted a field study that compared personality scores of actual job applicants on the NEO-PI-R to scores of job incumbents. It was found that job applicants scored significantly higher than job incumbents on all Big Five personality factor scores and on a measure of intentional response distortion. Thus it was concluded that applicants applying for actual jobs do in fact fake on personality tests. Similar evidence was presented by Robie, Zickar, and Schmit (2001), who compared 2 large field samples of sales manager applicants \( (N = 999) \) and sales manager incumbents \( (N = 769) \) from a large retail organization and found that the personality scale scores of applicants were approximately half a standard deviation higher than the scores of incumbents. More recently, a meta-analysis by Birkeland et al. (2006) compared applicant and incumbent scores on Big Five measures which were gathered from nearly 30 applied studies. The authors reported that applicants significantly outscored incumbents on all five of the personality factors. Four of these five effect sizes were statistically significant, with sizes ranging from .11 to .45 standard deviations. Other studies have also compared applicant and incumbent personality scores and found that applicants score significantly higher than incumbents (Hough, 1998b; Stewart, 1997; Van Iddekinge, Raymark, Eidson, & Putka, 2003; Weekley, Ployhart, & Harold, 2003).

An underlying weakness in many of the studies cited above, including those by Birkeland et al. (2006) and Robie et al. (2001), however, is that they used between-subjects designs to examine differences between applicants and incumbents. This design is plagued by the fact that although numerous differences (most notably motivation to fake) exist between job applicants and incumbents (Guion & Cranny, 1982), the between-subjects design treats these two groups as essentially equivalent. Furthermore, these
designs cannot utilize random assignment to the applicant and incumbent group and they
assume no subject by instruction interaction (Ones & Viswesvaran, 1998). For these
reasons, within-subject designs are seen as a better design for studying faking
(Lautenschlager, 1986). Although faking-based field studies using within-subjects designs
are uncommon in the literature, some recent studies provide valuable findings using this
type of design.

One such example is the recent study conducted by Ellingson, Sackett & Connelly
(2007), who examined a database of over 700 individuals who had completed the
California Psychological Inventory (CPI) twice through an assessment firm as part of
either an application or development process. Their unique database consisted of
individuals who had tested 1) once as applicants and once as incumbents, 2) both times as
applicants, or 3) both times as incumbents for development purposes. This allowed them
not only assess extent of faking by comparing Time 1 and Time 2 scores, but also to
isolate the extent to which changes in scores were attributable to intentional distortion by
accounting for the effects of time between administrations, feedback received on one’s
assessment, and practice effects. Their results indicated that the average change in scores
which could be attributed to actual faking was minimal – only .07 standard deviations.
The authors concluded that in applied settings, individuals do not engage in any
substantial faking, and that it has very little if any effect on hiring decisions. It should be
noted however, that in this study, the authors did not assess whether there were any
changes in the rank-order of applicants across test administrations. Thus, a portion of the
individuals may have scored significantly higher in the selection setting than the
developmental setting, which would increase their chances of being selected for a job.
Due to the fact that the authors only examined average changes in test scores across the full sample, however, this important outcome would not be observed or noted.

In a similar vein, Hogan et al. (2007) examined a dataset of 5,266 applicants who took a personality assessment in order to apply for a customer service position. They were all rejected and then re-applied six months later for the same position by completing the same assessment. By comparing mean scores on the two administrations, the authors were able to measure faking in a selection setting using a repeated measures design. Results indicated that only 5% of applicants raised their score on any scale during the second administration. Moreover, an equal percentage of applicants received lower scores on the second administration. They found that only 3 of the applicants increased their scores on all scales beyond a 95 percent confidence threshold, thus concluding that “faking on personality measures is not a significant problem in real-world selection settings.” Nonetheless, an important drawback of this study is that it was based entirely on a sample of applicants who failed the employment test the first time and as a result, were rejected for a position. In essence, the study only included individuals who had low propensity or ability to fake, thereby systematically removing the very applicants who would be most likely to fake and increase their score the most in order to increase their chance of being hired. Had the study included individuals who were successful in passing their first employment test, and then completed the test for development or research purposes, the results may have been much different.

In contrast to these findings, Griffith, Chemielowski, and Yoshita (2007) reported evidence of faking in a field setting using a repeated-measures design. Applicants for a customer service position completed a conscientiousness scale as part of the hiring
process. One month later, they were mailed the same measure \( N = 60 \) and were asked to complete it voluntarily under an honest condition. Participants scored significantly higher as applicants than they did under the honest condition. More importantly, they also found that the rank order of applicants was altered across these two test conditions such that the individuals who scored highest under the selection condition were not the same ones who scored highest under the research condition (Griffith et al., 2007). Given that organizations typically use fairly stringent selection ratios (Cascio, 1998) in order to hire the best possible applicants, the results of this study suggest that the company would have hired more customer service applicants who faked because these individuals tended to have the highest scores under selection condition.

A promising new study by Donovan, Dwight, and Hurtz (2003) assessed the prevalence of faking in applied settings without relying on any between- or within-subject comparisons. Using the randomized-response technique, which guarantees anonymity of all responses made by participants, the authors were able to obtain an estimate of the base rate of faking by actual entry-level job applicants. Subjects, who had all recently applied for various entry-level positions, were asked whether they had engaged in various faking behaviors, such as exaggerating one’s work experience or overemphasizing one’s positive attributes. Results revealed that a substantial portion of applicants reported engaging in various faking behaviors. The average prevalence rate of faking behaviors was 29%, but for some behaviors, it was as high as 77%. Nine of the 29 behaviors had a prevalence rate greater than 40% and 4 of the behaviors had a rate greater than 50%. The behavior with the greatest prevalence rate (76.7%) was “I exaggerated less than 10% of the information I provided”, whereas the behaviors with the lowest prevalence rate (0%) were “I
exaggerated about 75% or more of the information I provided” and “I claimed to have more education than I actually did.” Other behaviors which are highly relevant to selection scenarios were: “I exaggerated my work experience to make myself look more impressive than I really am” (44.7%) and “I gave responses on a test that were completely false or made-up.”

Moreover, since 75% of applicants reported that they felt the response technique truly guaranteed their anonymity and 68% indicated that the experimenters had no way of knowing which applicants admitted to faking, the authors concluded that the prevalence rates obtained from the applicants were reasonably accurate and representative of faking behavior in actual applicant situations. Most important, however, was that the prevalence rate of faking as measured by the randomized-response technique was over twice as high as the prevalence of faking which was obtained using a traditional questionnaire format, showcasing the power of this method to elicit more candid responses from respondents. These findings suggest that the aforementioned between-subjects studies may in fact be accurate in their reports that actual applicants do fake.

Although the debate over the prevalence of faking in actual selection settings continues to the present day, recent findings increasingly support the conclusion that faking is prevalent in applied settings and is therefore a critical issue to practitioners. However, this issue remains to be settled completely and it will only be resolved through continued faking research in actual selection settings. The best estimate of the prevalence of faking by applicants can only be gained by studying actual applicant behavior. Hence it is crucial for investigators to obtain data from actual organizations rather than from laboratory settings, which already have been used extensively in faking research.
Effects of Faking on Construct Validity

Most of the research concerning faking and construct validity has been based on the Five-Factor Model of personality (Borgatta, 1964; Costa & McCrae, 1992; Norman, 1961; Tupes & Christal, 1961). This model postulates that an individual’s personality can be categorized generally into five broad personality factors. These factors are: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. These factors, particularly Conscientiousness, have been found to predict job performance and training proficiency in applied settings (Barrick, Mount, & Judge, 2001; Barrick & Mount, 1991).

If faking does not affect construct validity, one would predict that the factor structure of personality tests administered to fakers should closely resemble the factor structure of tests administered to honest respondents. Hence, when the Five Factor Model is used, five factors should emerge when conducting a factor analysis for both faking and honest groups. However, a large body of evidence suggests that this is not actually the case.

Some of this evidence comes from laboratory studies where subjects were instructed to fake. One such study reported that when responses from subjects instructed to fake were compared with responses from subjects in an honest condition, exploratory factor analysis revealed that the two conditions had different factor structures and multitrait-multimethod (MTMM) matrices (Douglas et al., 1996). Frei, Griffith, Snell, McDaniel, and Douglas (1997) examined differences on a biodata scale between faking and honest groups using LISREL and found that the groups had a different number of latent factors which displayed different relationships among the observed variables. In a
study using a sample of ARMY personnel who were instructed to take the Assessment of Background and Life Experiences (ABLE; Hough et al., 1990) under both honest and faking instruction conditions, Ellingson, Sackett & Hough (1999) reported that in the faking condition, the factor structure of the ABLE was completely reduced to just one overall factor. Similar results using a factor structure other than the Big Five taxonomy have been reported as well (Holden & Jackson, 1981; Michaelis & Eysenck, 1971).

There appears to be a difference in factor structure of personality measures between applicant and student samples as well. Schmit and Ryan (1993), comparing the NEO-FFI responses of college undergraduates to those of a job applicant sample, found that the Five Factor model fit the responses of the student sample but did not fit the applicant responses. This discrepancy between the factor structure of the two samples was attributed to a sixth factor which emerged for the applicant sample only. Upon inspection, the factor had loadings on the conscientiousness, agreeableness and extraversion subscales, which led the authors to identify it as an “ideal employee factor.” Cellar, Miller, Doverspike, and Klawsky (1996) supported these findings by observing that a six factor solution fit NEO-PI responses from a sample of flight attendant trainees better than a five factor solution. Although in this study the data did not support the interpretation of this sixth factor specifically as an “ideal employee factor”, Cellar et al.’s (1996) results supports the finding that the factor structure for applicant personality measures differs from that of laboratory subjects.

Most compelling from an applied standpoint, however, is the finding that the personality factor structure of actual applicant data significantly differs from that of job incumbents. Van Iddekinge et al. (2003) reported significantly different factor structures
between customer service manager incumbents and applicants at a large retail organization. Weekley et al. (2003) also compared the factor structure of applicant responses to the structure of incumbent responses and found that, although the factor form was similar, the magnitude of the factor loadings significantly differed across the two groups. Even when partial invariance was tested, the fit of the statistical model was still poor, indicating that faking has detrimental effects on the construct validity of personality inventories. Thus faking can alter the factor structure of personality tests in laboratory settings and there is evidence that it may do the same in applied samples.

Some research findings, however, have indicated that faking does not significantly alter the construct validity or psychometric properties of personality inventories. Robie et al. (2001) analyzed data from 999 sales manager applicants and 796 sales manager incumbents at a large retail organization using Differential Item Functioning (DIF) analysis and found that although applicants tended to score higher on average, there were no psychometric measurement differences between the two groups. In other words, the items measured the same constructs regardless of whether they were administered to applicants or incumbents. This led the authors to conclude that applicant and incumbent personality scores do not differ in any practically significant manner. When Smith, Hanges, and Dickson (2001) compared the personality structure of responses from job applicants, job incumbents, and college undergraduates, the results indicated that the five-factor model of personality significantly fit each sample of data. This provided strong evidence that the factor structure of applicant personality test data is not significantly different from that of incumbents or students, suggesting that any effects of faking present in applicant settings are negligible. Ellingson, Smith, and Sackett (2001) categorized
individuals' personality scores into an honest response group and a faking group based on a social desirability scale (scores that were high on the social desirability scale were classified as faking scores) and found that the factor structure for both groups was highly similar. Therefore, they concluded that faking did not alter the factor structure of personality measures.

A meta-analysis by Ones, Viswesvaran, & Reiss (1996) confirmed similar findings regarding factor structure. Their results indicated that convergent validities, obtained by correlating each of the Big Five factor scores with the scores for same factor (e.g., Extraversion scores from one study with Extraversion scores from another study) across various studies and measures, remained stable after social desirability was partialled out of the correlations. Decreases in convergent validity coefficients only ranged from zero to .05, suggesting that faking has minimal, if any, effects on convergent validity of personality measures. The authors reported the similar findings for the discriminant validity coefficients, which should be low in order to demonstrate that each personality factor is unique and different from the other factors. Again, this was the case even after social desirability scores were partialled out of the correlations (Ones et al., 1996).

Moreover, in the study cited earlier by Hogan et al. (2007), the authors reported that the factor structure of the Hogan Personality Inventory (HPI), which is a Big Five-based personality measure, did not change across Time 1 and Time 2 administrations, which were both in an applicant setting.

Together, these studies do present a case that perhaps faking does not attenuate the construct validity of personality measures. Such findings are in direct opposition to those cited earlier, making it difficult to conclude exactly what the effects of faking on construct
validity are. Some have attributed the difference in findings to poor methods of statistical analysis, such as failure to account for multivariate normality in applied data sets, or the use of inappropriate fit indices in factor analysis, however, a review of the literature by Bradley and Hauenstein (2004) suggested that it is not likely that such methodological differences are the cause of the divergent findings. The authors state that other unidentified substantive factors are likely to be the cause of such differences across studies.

Although Bradley and Hauenstein (2004) fail to suggest why the findings from these studies are contradictory, it is my belief that a key factor is the way in which faking is conceptualized and measured. Two of the aforementioned studies (one being a meta-analysis of several studies) which found that faking did not alter construct validity used social desirability measures to classify respondents as fakers (Ellingson et al., 2001; Ones et al., 1996). This essentially equates faking with social desirability. Therefore, the findings from these two studies do not necessarily indicate that faking has no effect on construct validity; they simply indicate that social desirability has no effect, because that is how they operationalized faking. However, the crux of this proposed research is that it is problematic to equate faking with social desirability, and that research that uses social desirability measures cannot justify conclusions concerning faking. I will discuss this in greater detail in a later section of this review. Moreover, these studies only deal with construct validity. Even if faking has little effect on construct validity, faking could still have deleterious effects on criterion validity, which is of much greater importance to practitioners. Hence the main concern in faking research should be to assess the effects that faking has on the criterion validity of non-cognitive measures.
Effects of Faking on Criterion Validity – Faking Does Not Lower Validity

The effect of faking on criterion-related validity is complex and has been the subject of serious debate by researchers in the field. This complexity is due in part to conflicting findings on whether faking actually attenuates criterion validity. One body of research findings suggests that faking has little or no effect on the criterion-related validity of personality measures. Weekley et al. (2003) analyzed a sample of 2,989 job incumbents and 7,259 job applicants and found that although applicants had higher mean scores on three personality measures, criterion-related validities were equivalent across both samples, suggesting that any faking which may have occurred did not attenuate validity. Hough et al. (1990) reviewed criterion-related validities reported in studies from 1960 to 1984 and found that validity coefficients remained stable regardless of whether respondents engaged in “faking good” or “faking bad” response patterns.

Subsequent research in this area has also found that corrections for faking do not improve criterion validity. Christiansen et al. (1994) obtained measures on the 16PF from 495 assessment center applicants and compared them to subsequent performance measures. Correcting applicants’ scores for the two 16PF faking scales did not significantly improve criterion-related validity. Hough (1998) evaluated two strategies for dealing with applicant faking on a personality measure. One involved correcting applicants’ scores on a social desirability measure while the second consisted of removing applicants with extremely high scores on the social desirability measure from the applicant pool. Neither strategy was found to affect criterion-related validity. Barrick and Mount's (1996) research supports this claim as well. They tested the effects of impression management and self-deception, two types of response distortion (Paulhus, 1984), on the
predictive validity of the Big Five personality factors. Although subjects did engage in both of these types of response distortion, neither type lowered the predictive validities of the personality constructs. Furthermore, evidence from a recent simulation study concluded that under typical selection conditions, removing suspected fakers from an applicant pool has minimal effects on test validity (less than .10 standard deviations increase in mean performance ratings). The study also reported that there was even less impact on test validity when scores were statistically corrected for faking (Schmitt & Oswald, 2006). The authors noted, however, that although faking corrections or faker removal had minimal impacts on validity at an aggregate level, these procedures could have significant impact on selection decisions at the individual level.

Likewise meta-analytic studies finding that faking does not decrease predictive validity of personality or integrity tests have led some to claim that the data has "nailed the coffin of faking and social desirability shut" (Ones & Viswesvaran, 1996, p. 266). Their findings were based on the fact that partialing out social desirability from personality scores did not significantly improve validity coefficients across studies. The authors’ remarks emphasize that these research findings are convincing enough to conclude that social desirability and faking, in general, are not a threat to personality test validity, and the field should move on to other issues (Ones & Viswesvaran, 1998; Ones & Viswesvaran, 1996). Hogan et al. (2007) voiced a similar opinion based on their recent findings that applicants do not significantly increase their personality tests scores when re-applying for a job. However, this study did not examine faking effects on validity. I would like to note that in all of these studies, with the exception of Weekley et al. (2003) and the Schmitt & Oswald (2006) simulation, social desirability measures were used as
surrogates for faking and then statistically controlled for or partialled out of the predictor–criterion relationship. If social desirability does not fully capture faking behavior, however, such corrections could not be expected to improve criterion validity, and in this case, they did not.

**Effects of Faking on Criterion Validity – Faking Does Lower Validity**

On the opposing side of this debate are those who have demonstrated that faking does indeed lower criterion validity. Mueller-Hanson, Heggestad, and Thornton (2003) conducted a study in which participants completed the AMI-S, an achievement motivation measure, in either a control condition or an incentive condition, after which they also completed a 50-item performance task that consisted primarily of matching tasks, finding discrepancies, and simple problem-solving tasks, such as addition and pattern recognition tasks. In order to simulate an actual selection setting, participants in the incentive condition were informed that the investigators were interested in “people who are hard working, motivated, and conscientious” and that the test they were about to take measured these characteristics. They also were informed that only those who scored highest in these characteristics would be selected for the second half of the study, in which they would be eligible to win a cash prize of $20. Lastly, they were instructed to be honest in their responses because falsification could disqualify them. In contrast, subjects who were in the control condition were only instructed to be as honest as possible in their responses, and that their answers would be used only for research purposes and thus were completely anonymous. Results indicated that the validity of the AMI-S was lower for the incentive group ($r = .05, n.s.$) than for the control group ($r = .17, p < .05$). Although this difference
in correlation coefficients was not statistically significant, it does indicate that faking can actually lower validity coefficients.

Harold, McFarland, and Weekley (2003) investigated the effects of verifiability on the validity of biodata items in both an applicant and an incumbent sample. While verifiable biodata items were equally valid across both groups, this was not the case for non-verifiable biodata items. For these items, validity was significantly lower in the applicant sample than in the incumbent sample. The authors concluded that this difference was due to applicants faking on the non-verifiable items but not on verifiable items. Since applicants are more likely to distort their responses on items that cannot be verified, these items would be faked to a greater extent than items that could be verified, and applicants would have the motivation to do so. However, it was surmised that incumbents would have less motivation to fake and therefore, they responded similarly to both verifiable and non-verifiable items.

Komar, Brown, Komar & Robie (2008) conducted a simulation study that extended the aforementioned simulation by Schmitt & Oswald (2006). Rather than simply examine the effects of corrections on validity, they investigated how various parameters of faking would affect criterion validity. In contrast to Schmitt & Oswald’s (2006) conclusions, Komar et al. (2008) found that under certain conditions, faking can substantially reduce a test’s criterion validity. Critical variables such as the correlation between faking and performance, the proportion of fakers in the sample, and the magnitude of faking all exerted strong effects upon criterion validity. The faking-performance relationship had the strongest effect, potentially causing as much as a 74% loss in validity. The average validity loss across all parameter levels was 17%. The
authors also conducted utility analyses and indicated that on average, faking on an assessment would cost an organization over $520 per applicant, and in extreme cases, faking would cost an organization over $2400. Such findings clearly indicate that applicant faking can be very costly to any hiring organization, especially under high volume hiring conditions or when hiring highly paid candidates.

Dunnette, McCartney, Carlson, & Kirchner (1962) conducted one of the early studies on this issue and reported that criterion-related validity of personality scores for sales representative applicants who were instructed to fake was lower than for applicants instructed to respond honestly. Anderson et al. (1984) investigated the relationship between faking and clerical performance to assess effects on criterion-related validity. In fact, measures of faking had a significant negative correlation with performance on an external clerical task. Similarly, Pannone (1984) included a non-existent (fictitious) but seemingly job-related task on a biographical job questionnaire. Applicants who indicated prior experience performing such a task performed significantly worse on a written test of job knowledge. Moreover, when this group of fakers was removed from the analysis and validity coefficients were recomputed, validity coefficients increased from .42 to .55.

Similar results were found in a study which involved college students with previous work experience (Douglas, McDaniel, & Snell, 1996). Subjects' scores on a personality test and a biodata measure were correlated with previous supervisors' performance ratings. The results indicated that validities for those told to fake were much lower than for those told to respond honestly. Moreover, the study cited earlier by Cellar et al. (1996) lends partial support to this study as well. A full length version of the NEO-PI personality inventory and an adjective scale of bipolar personality items were administered to 423
flight attendant trainees. When validation of the two measures was attempted by comparing them to subsequent training success, the NEO-PI predicted success but the much more transparent adjective scale did not. Although no actual measure of faking was administered, it is possible to speculate that subjects distorted responses more on the adjective scale due to its high transparency (which is conducive to faking) and in turn, this led to poorer criterion-related validity in comparison with the NEO-PI, which has more subtle items. Subtle items do not make it clear to respondents what personality constructs are being measured, thus making them harder to fake.

Additional research has found similar results. Holden & Jackson (1981) found that the criterion validity of the Personality Research Form was lower when subjects were instructed to fake good or fake bad as opposed to when they were instructed to reply honestly. Topping & O’Gorman (1997) compared subjects’ responses on the NEO-FFI to personality ratings (on the same instrument) from judges who knew the subjects personally for at least 12 months and used the correlation between self and judge ratings as a measure of predictive validity of the NEO-FFI. Subjects were assigned to either an honest or a “fake good” condition. It was found that for 4 of the 5 Big Five factors measured by the NEO-FFI, validity was significantly lower for the faking group in comparison to the honest group. Along a similar vein, Worthington and Schlottmann (1986) found that the validity coefficients of two MMPI subscales in predicting a measure of nonconformity were only significant in an honest condition; they were not significant in a “fake good” or “fake bad” condition. In conclusion, there is a large body of evidence indicating that faking does lower criterion validity of various non-cognitive measures.
Problems With Validity Coefficients as Indicators of Faking Effects

Thus, in regards to criterion validity, the aforementioned research findings suggest that faking often lowers criterion validity, but not consistently. As discussed, some studies have found that faking does not significantly lower the criterion validity of non-cognitive measures (e.g., Ones et al., 1996). However, relying strictly upon attenuation of the validity coefficient for a whole applicant sample as an indicator of the effects of faking on validity may be misleading. Rosse et al. (1998) have pointed out several methodological reasons for why validity is not always affected even when faking does occur. First, the authors note that correlation coefficients are extremely robust in estimating a linear relationship between variables but this robustness results in a very low sensitivity to changes in rank order of particular ranges of a bivariate distribution. Therefore, when faking occurs, the rank order of applicant test scores may change at the top of the distribution but not in the remainder of the distribution, thus altering who is selected at the top of the distribution without lowering the magnitude of the test’s validity for the whole sample (this issue will be discussed in greater detail in the following section).

Second, applicant scores on faking measures tend to be negatively skewed (moderate to high levels of impression management), causing range restriction in the predictor. This makes it difficult to detect relationships between faking and other variables. Third, low selection ratios compound the problem posed by a negative skew, resulting in further range restriction. Since low selection ratios are desirable and typical in actual organizational settings, this is a common problem in applied data sets. Finally, the suppression effects of faking on criterion validity are extremely hard to detect when
validity coefficients are low to moderate in magnitude, which is almost exclusively the case for personality measures in organizational settings. It is possible that these methodological factors prevented some researchers from detecting a detrimental effect of faking on criterion validity in their data sets (e.g., Hough et al., 1990; Weekley et al., 2003). Thus, it is important to look beyond just the correlation coefficient when assessing whether or not faking decays criterion validity.

In addition to the methodological factors that may obscure the effect that faking has on the validity of non-cognitive measures, the failure to account for variance due to individual differences in faking may also be the cause of so many discrepant findings in this area. Lautenschlager (1986) addressed this issue by stating that if all applicants fake to the same extent, then faking does not change rank orders of applicants and therefore criterion validity should be unaffected. However, if there are individual differences in the extent to which applicants fake, then faking can affect criterion validity. Research findings support the notion that such individual differences do exist. For example, individuals who are low in integrity, low in conscientiousness, and high in neuroticism tend to fake to a greater extent on personality, biodata, and integrity tests. In addition, these differences tend to be stable across the three test types (McFarland & Ryan, 2000). Similarly, Griffith, English, Yoshita, Gujar, Monnot, Malm, and Graseck (2004) found that individual differences in integrity and locus of control are also related to applicant faking behavior.

Hence it seems unwise to assume that all individuals fake to the same extent and in the same fashion. Instead, research suggests that applicants differ in these behaviors, making it possible for faking to change the rank ordering of applicants on a non-cognitive
measure. If this is the case, conclusions from studies that partial out faking or social desirability from personality test scores may be erroneous. Partialing out the contamination of faking incorrectly assumes that there is no variance on the faking itself (Bowen et al., 2002) despite the aforementioned evidence that there is substantial variance in faking (Griffith et al., 2004; McFarland & Ryan, 2000). This methodological practice could be the reason that several studies fail to find decay in validity coefficients when faking is controlled statistically. However, this only holds true for the studies which used this methodology (e.g., Barrick and Mount, 1996; Ones & Viswesvaran, 1996). It would not apply to studies concluding similar findings based simply on comparisons of applicant and incumbent samples without an actual measure of faking (e.g., Weekley et al., 2003).

In addition to the failure to account for individual differences in faking, a methodological flaw in studies that partial out social desirability is the underlying assumption that faking equals social desirability. In the context of this literature review, this methodological flaw is the most important and will be thoroughly discussed in an upcoming section. Just as I discussed in the earlier section on construct validity, the same theoretical dilemma affects criterion validity analyses. Simply put, if a researcher statistically partials out social desirability from the predictor – performance relationship and this fails to increase the correlation coefficient, it does not necessarily imply that faking does not affect validity. It simply implies that social desirability does not affect criterion validity. Unfortunately, the majority of existing applied faking studies have made the former assumption. Nonetheless, I propose that until there is fairly conclusive evidence that social desirability alone can account for all of the variance in faking behavior, this methodological approach is inappropriate.
To summarize on the issue of whether correlation coefficients are sensitive to faking, the existing applied studies reviewed here generally have used one of two strategies to test this hypothesis: 1) assessing whether the criterion validity of applicants’ scores is lower than the validity of job incumbents’ scores, and 2) testing whether criterion validity is improved after statistically controlling for social desirability. Both strategies are subject to methodological constraints which could preclude researchers from detecting the effects of faking on criterion validity. In the case of straight comparisons between applicants and incumbents, the methodological issues outlined by Rosse et al. (1998), such as insensitivity of correlations to rank-order changes and range restriction, present the problem. As for the studies using statistical control for social desirability, these studies suffer from not only from some of the issues outlined by Rosse et al. (1998), but also from the assumption that social desirability is an adequate measure of faking and from the failure to account for individual differences in faking. For these reasons, researchers using these methodologies may have prematurely suggested that faking does not attenuate the predictive utility of non-cognitive measures. Hence, it is necessary to look beyond these approaches and instead investigate whether faking adversely affects final hiring decisions, rather than only assessing its effect on validity coefficients. One such approach is discussed in the following section.

_Fakers “Rise to the Top” Regardless of Stable Validities_

One way to assess the impact of faking on hiring decisions is to examine whether applicants who fake are the same applicants who tend to be hired. Despite the fact that validity coefficients may remain intact when applicants fake on non-cognitive measures, research has found that applicants who fake will still “rise to the top” of the applicant pool
because they raise their test scores inordinately over the scores of non-fakers. Although this does not guarantee that a person who is low on a trait will outscore a person who is naturally high on that given trait, this process does become a problem in the upper portions of the score distribution. Those who score in the highest percentiles because they are naturally high on a trait cannot score any higher due to the nature of the measure being used. Therefore, this range restriction causes a “ceiling effect” on their scores, which prevents these non-fakers from increasing their scores, thereby allowing fakers (who would have scored much lower if they had answered honestly) to significantly increase their scores to the top portions of the score distribution and “catch up” with non-fakers. This is how faking actually leads to rank-order changes in non-cognitive test scores (Lautenschlager, 1986; Nunnally, 1978). This process, in turn, adversely affects potential hiring decisions in a top-down selection process where an organization hires applicants with the highest scores on a given non-cognitive test score.

Using a sample of actual job applicants for various entry-level positions at a ski resort, Rosse et al. (1998) found that with selection ratios of 25% or lower, the average level or faking for applicants hired was at least one standard deviation above the mean. At a selection ratio of 5%, faking scores were nearly two standard deviations above the mean. In more practical terms, if only the top 5% of applicants are hired, 7 of the 8 people hired would have extremely high scores on a faking measure. Similarly, if the top 10% are hired, over half of the new hires would have extremely high faking scores.

Previous research using computer simulations has led to the same conclusions about the effects of faking on the rank-ordering of job applicants. For example, Zickar, Rosse, and Levin (1996) and Douglas et al. (1996) conducted Monte Carlo studies which
simulated applicant faking on non-cognitive measures. Both studies demonstrated that even with relatively few fakers in a sample, the top end of the distribution can contain a high percentage of fakers. In Zickar et al.'s (1996) study, faking had a noticeable effect on who would be hired although it had no effect on predictive validity. For example, Zickar et al. (1996) concluded that if ten percent of the subjects fake, five to six of the top ten subjects will most likely be fakers. It should be noted that in selection settings, these top ten subjects would most likely be selected for the job of interest. Moreover, in the Douglas et al. (1996) study, the validity coefficient calculated for the entire sample was .20 whereas the validity was near zero if calculated only for those with the top ten scores. In the Ones et al. (1996) study, validity coefficients were based on the entire sample of applicants, not just the ones who were hired or selected. If the Ones et al. (1996) meta-analysis had calculated validities based only on those selected (at the top of the distribution), validity coefficients may have been dramatically lower (Miller, 2000).

It is possible to argue that the results of the Douglas et al. (1996) study described above were attributable to range restriction in the predictor scores, whereby validity coefficients based only on the top scores were highly attenuated. However, the Mueller-Hanson et al. (2003) study cited earlier provides the same findings in a manner which circumvents this issue. As previously stated, one group was in an honest response condition while the other group was in an incentive condition. When the incentive group was divided into top and low thirds based on AMI-S (measure of achievement motivation) scores, validity was significantly higher for the lower third \( r = .45, p < .05 \) than for the top third \( r = .07, ns \). However, this was not the case in the control group, as there was no significant difference in validity coefficients when the top and low scorers were
compared. Thus the authors suggested that in the applicant condition, fakers were able to rise to the top of the distribution, changing the rank-order of the upper distribution and thus decaying the validity of the predictor for only these applicants. The lower part of the distribution, in contrast, did not consist of many fakers, thus the validity of the predictor for applicants in this score range was maintained. As for subjects in the honest condition, there was no difference in validity between upper and lower thirds of the score range because there was no inherent motive to fake. Due to the fact that the study compared validities for the upper and lower thirds of the score range (as opposed the comparing the upper and full ranges), the difference in validities are not necessarily attributable to range restriction effects, as in the case of the Douglas et al. (1996) study.

Even more striking, however, is the finding by Mueller-Hanson et al. (2003) that as the selection ratio decreases (e.g., as an organization becomes more selective) the percentage of fakers selected disproportionately increases. In fact at selection ratios of .50, .25, .20, .15, and .10, which most adequately represent actual selection practice, the number of honest respondents that were selected was significantly and consistently lower than the percentage of honest respondents in the full sample. This was not the case for selection ratios of .60, .70, .80, or .90, however, indicating that fakers indeed rise to the top of the distribution and greatly improve their chances of being selected on the basis of a non-cognitive test score.

Additional research has documented the same effect. Hough (1998) also found that fakers rose to the top when she administered a personality inventory and a social desirability measure to a sample of actual police applicants and incumbents. Any police applicants whose social desirability score was 95% higher than that of police incumbents
were classified as fakers. When results were analyzed and the top 10% of the applicant scores were selected, only 18% of the total applicant sample consisted of fakers. However, 46% of the selected applicants were fakers. A similar study which simulated a selection setting for safety force personnel positions found even more striking results – participants who received coaching on how to fake a conscientiousness, or were instructed to fake on it, made up 100% of the top 10% of those selected based on this test. Of those who were instructed to respond honestly, none were selected based on a top-down selection situation where the top 10% were hired. Such a hiring ratio is very typical of selection for these types of positions (Miller & Barrett, 2008).

Thus it is clear that faking does lower the predictive validity of personality measures because it affects who is selected at the final stage of the selection process, regardless of what the validity coefficient of the whole sample is. Researchers could possibly calculate validity coefficients based only on those applicants who are selected in order to better assess the effects of faking on validity of personality measures. However, this would introduce range restriction, which would give an artificially low estimate of validity. Nevertheless, such research could still lead to valuable findings in this area of selection.

Taken together, the findings from these studies highlight why it is essential to look beyond simple correlation coefficients and instead examine actual hiring decision processes when studying how faking affects validity. In fact, some researchers have already suggested that criterion validity may not be a good indicator of a selection measure’s utility because it is insensitive to deleterious effects of faking (Douglas et al., 1996; Rosse et al., 1998; Zickar, et al., 1996). On the other hand, some researchers have strongly criticized this argument on grounds that criterion validity is a crucial index
of how well a predictor functions and it has guided the science of industrial and personnel psychology research for decades (Ones & Viswesvaran, 1998). Moreover, criterion validity is the most feasible validation strategy for non-cognitive measures. In fact, these researchers go so far as to state that shifting the focus away from criterion validity when evaluating the predictive utility of a test is “equivalent to shifting into a clinical approach in personnel selection.” (Ones & Viswesvaran, 1998, p. 256). Hence the call to look beyond criterion validity in faking and selection research is not without its critics, but in my opinion, it is a valuable and necessary suggestion in the field of applicant faking research because it leads to alternative methods (such as examining whether fakers rise to the top) which unveil how faking compromises the validity of actual selection situations whereas relying on validity coefficients would not consistently reveal these important effects.

Faking is Not Equivalent to Social Desirability – Theory

In order to further examine why faking does not consistently lower validity, it is crucial to consider the way in which we conceptualize and operationalize faking behavior. These findings may be largely due to the fact that most of the earlier research on faking defined and measured faking as social desirability. Examples of these measures are well-known validity, or lie, scales such as the Minnesota Multiphasic Personality Inventory (MMPI) F and K Scales, the CPI Good Impression Scale, and the 16PF Faking Good and Faking Bad Scales. Such scales are designed to identify this particular type of response distortion, social desirability, because the theory underlying their construction equates faking with social desirability. However, various researchers do not believe social desirability represents faking. Instead, they define social desirability as an aspect of a
healthy personality (*e.g.*, Hogan, Hogan, & Roberts, 1996; McCrae & Costa, 1983). These researchers argue that removing social desirability from personality test responses decreases their validity.

As I have previously stated, this presents a problem in that if social desirability does not constitute faking, then these types of faking measures may be ineffective in detecting faking. In fact, this is precisely the case. Various researchers have discussed and shown that faking is much more than just social desirability and brought forth evidence that traditional social desirability scales are not necessarily the most effective measures of faking.

Social desirability refers to the general tendency to distort self-reports in an overall favorable direction (Furnham, 1986). A similar definition of social desirability, proposed earlier by Edwards (1957, p.32), conveys the same general meaning. He described it as "a tendency to attribute to oneself personality statements with socially desirable values and to reject those with socially undesirable values." Stated differently by others, social desirability refers to a need to obtain approval by responding in a culturally appropriate and acceptable manner (Marlowe & Crowne, 1961). Hence, the social desirability construct can be defined best in terms of people's attempts to make a general, all-encompassing good impression in self-reports without intentionally creating any specific or unique impression. This is one of two major characteristics of social desirability.

The second aspect of the social desirability construct deals with intention, which gives it a two-factor structure. It has been shown that social desirability can actually be divided into 1) self-deception and 2) impression management (Paulhus, 1984; 1991b). Self-deception can be defined as an unconscious enhancement of the ego which leads one to
form overly positive beliefs about one’s self-concept. Thus, this factor actually represents a true dimension of a person’s personality and should be treated as true variance in personality. It has been shown to correlate highly with facets of adjustment, such as self-esteem, depression, and neuroticism (Paulhus, 1991b). Thus the self-deception factor in social desirability is unconscious and unintentional.

The second factor in social desirability, impression management, is different in this regard because it refers to the conscious, deliberate falsification or distortion of one’s true personality in order to create a positive impression. This factor does not represent true variance in a person’s personality and thus introduces error variance. The key aspect of this distinction that is relevant to the present review is that the self-deception is unintentional response distortion whereas impression management is intentional response distortion. Whenever response distortion is intentional and deliberate, such as in the case of impression management, the term “faking” applies. This is one of the major differences between social desirability and faking. Since only one part, or type, of social desirability is intentional – impression management, then only this type should be considered faking.

As stated previously, however, impression management is a faking strategy that attempts to create positive, but general, impression. It does not lead to any specific type of impression, but attempts to make the respondent “look good” on all dimensions of a measure. Therefore impression management represents only one way to fake; it is not the only way in which a person can engage in deliberate response distortion. It is my belief that the inconsistencies in the literature regarding the effects of faking on construct and, more importantly, criterion validity of non-cognitive measures is due to the common
assumption that faking is equivalent to either: 1) the whole social desirability construct, or 2) the impression management factor in social desirability. Faking is not entirely equivalent to either of these because it is deliberate and it can involve creating a specific impression, not just an overall, general impression. Therefore, controlling statistically only for social desirability measures, or components of social desirability measures will not completely control for faking because it may not capture all of the ways in which people can distort their response. This could explain why in the Ones et al. (1996) meta-analysis, it was found that faking did not attenuate validity. This finding could be due in part to the fact that there was no distinction made between the two qualitatively distinct dimensions of social desirability – impression management and self-deception. This is an important distinction because, as stated earlier, it distinguishes between intentional and unintentional response distortion.

**Faking is Not Equivalent to Social Desirability – Research Findings**

Although the majority of past research on faking has measured the construct with social desirability measures (e.g., Hough et al., 1990; Ones et al., 1996), a growing body of evidence suggests that these two constructs are not the same thing. One initial problem is that different social desirability scales, which purport to measure the same construct, do not tend to correlate highly with each other, which demonstrates low convergent validity (Paulhus, 1991b). This suggests that the construct of social desirability itself needs to be more narrowly defined. Another issue, described earlier, is that social desirability scales have been found to correlate highly with personality traits such as adjustment, conscientiousness, emotional stability, and integrity (Ones et al, 1996; Zerbe & Paulhus, 1987). Although this most likely is due to the relation between the self-deception factor
in social desirability and personality, it still indicates it is problematic to use the whole social desirability construct as a surrogate for faking because it is partly related to true personality variance.

Some research even indicates that social desirability is not a strong predictor of faking, which ultimately defeats the purpose of using social desirability scales to measure faking. Griffith et al. (2004) reported that neither the self-deception nor the impression management subscales of the BIDR, a popular social desirability measure, significantly predicted faking on the NEO-FFI. Faking in this study was defined as scoring above the upper bound of a 95% confidence interval established around a subject’s score in an honest response condition. This condition followed an initial faking condition in which all subjects first were led to believe they were applying for an actual clerical position which paid $18 per hour. Hence, self-deception and impression management, which have been traditionally associated with faking and have been used to detect faking so often, may not be strongly associated with faking behavior after all.

Despite the findings from this line of research, many studies continue to conceptualize faking simply as social desirability, and measure it only with social desirability scales. Upon finding that social desirability does not affect psychometric properties of the scales to any significant extent, these studies then infer that faking has no effect on personality measures either. However, if one takes into account the aforementioned problems of low convergent validity, shared variance with true personality, and weak relationships with faking behavior, it can be argued that the reason these studies have found no effects of faking is because perhaps they were not properly measuring faking.
Two previous studies serve as a useful example to this argument. In both studies, faking was measured with social desirability scales. A closer look at these studies illustrates why the use of social desirability scales to measure faking may have failed to fully capture faking behavior and led to the erroneous conclusion that faking had effect on personality scores. In one study, Ellingson et al., (2001) categorized individuals' personality scores into an honest response group and a faking group based on a social desirability scale (scores that were high on the social desirability scale were classified as faking scores) and found that the factor structure for both groups was the same. Therefore, they concluded that faking did not alter the factor structure of personality measures. In another study, the "Big Five" factor structure was imposed on three different sets of scores on the HPI obtained from students, job applicants, and job incumbents. Results indicated that the number of factors underlying the HPI scales did not change across all three sets of scores (Smith et al., 2001).

One important aspect of both the Ellingson et al., (2001) and the Smith, Hanges, and Dickson (2001) studies is that neither investigated factor structures within each type of job. The majority of the subjects used in the two studies were actual job applicants who were applying for a wide variety of jobs. However, both studies collapsed subjects across all job categories, so that the type of job being applied for was not accounted for in the analyses. This would allow fakers to simply fake on items that seemed to be relevant to the job they were applying for, instead of faking on all possible items. If this were the case, it would be difficult to detect faking because only a portion of the total items would be faked (those relevant to the job). An earlier study by Ellingson et al., (1999) further illustrates this argument. Their data was obtained by instructing Army personnel to
complete the ABLE first under an honest condition, then under conditions to respond in a way which would ensure that the Army selected them (i.e., in an overall, socially desirable way). They found that correcting each individual’s ABLE score in the faking-condition for social desirability resulted in scores that were very different from their honest ABLE scores. In other words, the factor structure of the ABLE dissolved into one overall social desirability factor. In addition, they found that corrected faking scores on the ABLE had very low construct validity due to their low correlations with the honest scores. Because this study instructed subjects to fake in an overall, socially desirable manner, subjects most likely faked on all possible items, thereby eroding the factor structure that the test was designed to measure. Thus it would be much easier to detect such extensive mean score differences across the honest and faking conditions under these circumstances, whereas in the prior two studies (Ellingson et al., 2001, and Smith, Hanges, & Dickson, 2001) faking may have been very subtle and difficult to detect because subjects had a specific job in mind when they applied. This would allow them to fake on far fewer items without being detected by the social desirability measures because social desirability measures are not designed to detect such specific faking. Moreover, some researchers cite anecdotal reports of actual applicants suggesting that social desirability scales themselves may be susceptible to faking because applicants fear that their responses might not be believable if they appear to make themselves look “too perfect” (Barrett, 2001).

Research using methodology based on Item Response Theory (IRT; Drasgow & Hulin, 1990) further weakens the argument that social desirability can capture applicant faking. Stark, Chernyshenko, Chan, Lee, and Drasgow (2001) found that items from all
subscales of the 16PF personality questionnaire functioned differently across applicant and non-applicant samples to a substantial degree, suggesting that applicants did in fact fake on the 16PF. Interestingly, the impression management scale used in the study to measure faking exhibited differences in item functioning as well, which is tantamount to applicants faking the faking scale, as suggested by Miller & Barrett (2001). Therefore, the same impression management scale measured different constructs, depending on whether it was used on an applicant or a non-applicant sample. Stark et al. (2001) concluded that studies classifying individuals as honest or faking respondents on the basis of social desirability scales (e.g., Ellingson et al., 2001) are therefore problematic. The problem is that these studies classify applicants as either fakers or honest respondents based on how highly they score on social desirability scales. However, the social desirability scales may not be sensitive enough to detect faking because faking is more than just social desirability.

*A Better Description of Applicant Faking – Job Desirability*

It follows that if faking is not comprised solely of social desirability, perhaps another form of response distortion accounts for the variance in faking. Ones et al. (1996) alluded to this notion when they mentioned the possibility of a response distortion bias toward the specific applicant situation. Kluger and Colella (1993) referred to this as "job-specific bias" or "job desirability." They defined this as the tendency for individuals to present themselves as having characteristics deemed desirable for the specific job for which they are applying for. Job desirability is different than social desirability in the sense that applicants would present themselves in a manner that would give them the best
chance at being selected for a specific job, whereas they would not necessarily present themselves in a socially desirable way.

To further illustrate the difference between job desirability and social desirability, an example of job desirability would be that of an applicant for a law enforcement position who falsely responds that he is highly aggressive and dominant because he feels these qualities are necessary for being a good police officer. Even though he might feel as though these traits might be seen as socially undesirable in an everyday sense, he would sacrifice giving the socially desirable response (being low on aggression and dominance) in order to give the response which will give him the best chance to be hired (being high on both of these traits) because he believes that these traits are desirable in a law enforcement officer.

Conversely, an example of social desirability\(^1\) would be that of a law enforcement applicant who falsely responds that he is low on aggression and dominance because he feels that these qualities are not socially desirable in an everyday sense. If he responded in a socially desirable manner on all personality items, this applicant would attempt to score very highly on any traits which society deems desirable, while scoring very low on any traits which society deems undesirable - regardless of whether the job he is applying for requires that trait. Socially desirable responding is thus general in the sense that it is directed at every personality trait which the person can distinguish on the test they take. In contrast, job desirable responding would not necessarily be reflected on every

---

\(^1\) It should be noted that in this particular example, when I use the term “social desirability,” I am referring solely to the impression management factor of social desirability because this represents intentional response distortion. I have used the more general term “social desirability” in this instance because the extant literature usually uses this broader term to describe deliberate faking, even though it may refer to the whole construct or just the impression management factor. More importantly it focuses the reader’s attention on how job desirability and social desirability differ in the aspect of generality versus specificity.
personality trait; people would only fake on traits deemed relevant for the job they are applying for.

Given that job desirable responding would result in responses which are specific to a certain job, it implies that applicants would respond differently for different jobs. In fact, many studies have documented this specific finding. Wesman (1952) found that the same individuals gave different profiles when they were instructed to answer as if applying for a sales position versus when applying for a librarian position. Stanley and Stokes (1999) found that subjects provided very different personality profiles when given different job descriptions and then asked to fake applying for the positions of Air Force officer, graphic artist, and marketing executive. Mahar, Cologon and Duck (1995) reported that subjects who were told to give the best general impression themselves responded quite differently than those told to present themselves as suitable for a psychiatric nurse position. Similar results were found by Kluger, Reilly, and Russell (1991), who found that subjects who were given information about a specific target job biased their responses according to the job specifications, resulting in different scores than those who did not receive any information about the target job. And in yet another study, Velicer and Weiner (1975) reported that the same subject gave different responses when instructed to either respond honestly, to portray an ideal self, or respond as if applying for library or sales positions.

The Measurement of Job Desirability

Although the concept of job desirability as a form of faking is relatively new in the literature, a body of work is beginning to accumulate which has described and measured this particular type of response bias in personnel selection. Other terms that have been
used are “inflation bias” (Anderson et al., 1984) and “job-relevant pattern faking” (Raymark, Shilobod, & Steffensmeier, 2004). Using various methods, these studies have consistently found that job applicants often fake their responses in a manner which is very specific and tailored to the job for which they are applying, indicating that social desirability may only measure one aspect of faking whereas job desirability captures a different aspect.

Anderson et al. (1984) found evidence for a job-specific response bias when they asked job applicants for a clerical position to answer a questionnaire which contained embedded bogus tasks which superficially resembled job-relevant tasks but were in fact non-existent. An example item is “I have experience matrixing solvency files,” which indicates faking if the respondent replies that he or she has done this task in the past. Surprisingly, nearly one-half of the applicants faked at least one of these items by claiming they had experience performing such non-existent tasks (Anderson et al., 1984). Pannone's (1984) study was very similar in that he found that over one-third of applicants for an electrician position claimed they knew how to operate a piece of non-existent equipment on a questionnaire testing job-related knowledge.

Using a similar “bogus-item” approach, Carroll, Jones, and Sulsky (2004) developed a test containing bogus items about popular culture which was used in a simulated application setting to select a candidate for an “Events Coordinator” at a college campus. The test contained 75 names, 25 of which were fictitious, and participants were instructed to rate their familiarity with each name. One group of participants was informed that the ideal job candidate for this position would have knowledge of popular culture and familiarity with various performers and literary figures. This group was also
told that the candidate who scored highest on this questionnaire would win a cash prize.

A second group in a control condition also completed the questionnaire but was not informed about the job description or the cash prize, but simply was instructed to respond as honestly as possible. Results indicated that the bogus item test displayed adequate construct validity as a measure of faking as evidenced by significant correlations with measures of impression management and self-reported admission to faking and non-significant correlations with a self-deception measure. More importantly, the test was successful at identifying fakers; subjects in the incentive condition scored significantly higher on the bogus item test than subjects in the control condition.

Raymark et al. (2004) adopted a different and more complex strategy in order to detect job desirable faking. Using only facets from the NEO-PI-R (Costa & McCrae, 1992) personality inventory, they reviewed a vast number of specific personality facets (subcomponents of the larger “factors) and theoretically linked them to personality requirements of several jobs which had very distinct job descriptions. This effort resulted in a set of six different jobs which required candidates who possessed substantively different patterns of the Big Five personality factors to be successful at each job. It was ensured that each job had an optimal personality profile in which at least one of the personality-performance relations was in the opposite direction as would be found in a socially desirable response set (the previous example of a law enforcement applicant responding high on “aggression” illustrates this point; the job desirable response for this job would be “highly aggressive” whereas the socially desirable response would be “not aggressive”). Raymark et al. (2004) reported that faking was very context-specific in that people produced profiles that were consistent with the requirements of the job. More
importantly, subjects raised their scores on some facets but lowered their scores on other facets even if, theoretically, the socially desirable response would have been to have a high score on that particular facet.

Similar results were reported earlier by Mudgett (2000), who experimentally manipulated the type of job that participants were instructed to fake for. He found that the desirability of different personality traits varied for different jobs. Moreover, results indicated that participants did not respond to the personality measures in a uniformly socially desirable manner, but in a manner which was desirable for the target job. This involved changes in magnitude (e.g., endorsing an item as “very frequently” rather than “frequently”) and direction (e.g., endorsing an item as “frequently” rather than “sometimes”) of faking according to job type. Both these studies provide excellent examples of a situation in which using a social desirability measure to detect applicant faking would be inadequate because of the complex relationship between faking and the type of job in question. Whether social desirability or job desirability measures are more effective in detecting fakers is the topic of the next section.

Job Desirability Measures: More Sensitive to Faking than Social Desirability Measures

Although the Raymark et al. (2004) and the Mudgett (2000) study cited above did suggest that job desirability may offer a more effective approach to detecting fakers than social desirability, these studies did not directly test this hypothesis. In fact, very little research has specifically examined whether some type of job desirability measure would be more sensitive to faking than a social desirability measure. One previous study (Kluger & Colella, 1993) sought to answer this particular question, albeit using a biodata test rather than a personality test, administering a biodata test to 429 individuals applying
for a nurse's assistant position. Faking was operationalized as responses to items on the biodata measure that were on extreme ends of a Likert-type scale. Measures of both social desirability and job desirability were included. Using hierarchical regression, job desirability had an incremental effect above and beyond social desirability in predicting the extreme responses, whereas social desirability did not have an incremental effect above and beyond job desirability. The authors concluded that job desirability was a better predictor of faking biases than social desirability.

Subsequent research in this area also has found that job desirability measures detect faking more effectively. To investigate this issue, Miller and Tristan (2002) conducted a study where a sample of 599 undergraduate students underwent one trial in which they completed the NEO-FFI personality inventory under instructions to answer in a manner that would give them the best possible chance of being hired for a law enforcement position. In a later trial, they completed the same measure under instructions to answer as honestly as possible. Faking was operationalized as the difference in NEO-FFI conscientiousness scores in the first trial minus their scores in the second trial. Subjects completed social desirability and job desirability measures in each trial. Results indicated that although the social desirability measures were related to standardized mean differences (Cohen's D) ranging from 0.05 to 1.63 standard deviations, the job desirability measure was much more sensitive to faking, resulting in a Cohen's D of 2.63 standard deviations. Thus, a higher Cohen's D, or effect size, in score differences between faking and honest conditions was found for job desirability scores than for social desirability scores. An earlier study by Miller and Tristan (2002) using similar design and measures
led to the same finding, indicating that job desirability measures are more sensitive to faking than social desirability measures.

Nonetheless, some findings indicate that social desirability measures are quite sensitive to faking on personality tests. Viswesvaran and Ones’s (1999) meta-analysis on the fakability estimates of personality measures found that, on average, social desirability scales had higher mean score differences between honest and faking conditions than any of the Big Five factor scales as evidenced by the highest effect sizes between test conditions. This was the case regardless of whether the studies used a between- or within-subjects design. The authors concluded from this review that social desirability scales are very useful in the detection of faking. However, it should be noted that in this study, the sensitivity of social desirability measures to faking was only compared to that of Big Five personality scales, which are not specifically suited for the detection of faking. Had a comparison been made against another type of faking measure, such as a job desirability scale, results would have probably been different.

Although the aforementioned findings certainly suggest that job desirable responding more accurately captures how an applicant may fake a test, it does not necessarily imply that applicants do not engage in any socially desirable response bias at all. It is certainly feasible that some combination of these two types of faking may occur as a person attempts to convey an optimal impression to an employer. Moreover, the studies which compared effect sizes of the two types of tests across faking and honest conditions (Kluger & Colella, 1993; Miller & Tristan, 2002) revealed that although job desirability scales displayed the largest effect sizes when they were faked by respondents,
social desirability measures were also faked consistently. Thus, at least some degree of socially desirable responding does seem to go on as well.

In fact, some research supports this argument, implying that applicants may utilize both job and social desirability strategies when they fake on personality tests, depending on how relevant the personality items appear to be for the target job. Burkrant (2001) conducted a study that examined subjects’ profiles on personality items that were relevant to a target job and on items that were irrelevant to the target job that they were instructed to fake for. It was hypothesized that if applying for a journalist position, for example, subjects would fake being more agreeable, extraverted, and less conscientious than if applying for a Certified Public Assistant (CPA) position, which instead would lead to subjects faking higher conscientiousness, and less agreeableness, and extraversion. Results revealed an interesting pattern; for items that measured job-relevant traits, subjects faked in a job-desirable manner, which was consistent with the hypotheses. However, for items that were irrelevant to the target job, subjects tended to answer in a socially desirable manner instead, as evidenced by high scores on all possible traits – not only the ones relevant to the job.

If Burkrant’s (2001) findings are applied to previous research, this would imply that in the Ellingson et al., (1999) Army study, subjects would have faked on social desirability because there was no specific target job. Hence, subjects most likely faked on all possible items, which altered the factor structure of the personality tests extensively. Such sizeable factor structure changes in the personality measure would be easy to detect. But, in the Ellingson et al., (2001) and Smith, Hanges, and Dickson (2001) studies, subjects would have faked on job desirability because they were applying for specific
jobs. This job desirability faking strategy would result in changes that were much more subtle, thus the social desirability measures used in those studies to detect the faking was not sensitive enough to detect these subtle changes in the psychometric properties of the personality measures that were used.

Above all of the problems discussed earlier on this matter, it is this lack of measurement sensitivity (to faking) which most prompts me to propose the use of job desirability measures, rather than social desirability measures, to identify applicant faking on non-cognitive measures. There is ample evidence at this point to conclude that applicants can fake in ways which are not entirely captured by the social desirability construct. Job desirability has been found to successfully identify fakers and it is more sensitive to faking than social desirability. Therefore, the use of social desirability scales alone as a means for identifying fakers seems unwarranted. Job desirability measures can tap into subtle and job-specific faking variance in ways which social desirability measures simply cannot.

Aside from the research body supporting it, this argument is rational in a theoretical sense because each applicant is different and each job description is different, thus faking behavior depends on the job and the applicant. A measure of general social desirability or impression management, which would be the same regardless of the job, would be unable to measure important aspects of each unique selection situation. In contrast, a job desirability measure will always be specifically developed to be highly job specific. I believe that this is the reason why so much of the research on the effects of faking and validity is not more conclusive. If more studies had measured faking with a job-specific measure, I believe construct and criterion validity would have deteriorated
more consistently as a consequence of this type of faking. Then, such findings would concur with the evidence from studies investigating the “rise to the top” effect (e.g., Rosse et al., 1998), which found that those who were most dishonest in their responses were most likely to be hired.

**Conclusion**

In conclusion, a review of the literature identifies five main issues. First, there is no longer any question as to whether most non-cognitive measures can be faked; evidence consistently shows they can be faked to a substantial extent. Second, there is mounting evidence that faking is prevalent in applied settings, although there is still some debate surrounding this issue. Third, research has shown that some methods can help deter faking on non-cognitive measures. Fourth, faking has been shown to lower both construct and criterion validity, although this effect is not always found. Regardless, it is clear that faking can still have strong effects on hiring decisions even when criterion validity is statistically unaffected. Fifth, research suggests that faking consists of more than just the social desirability construct. Job desirability is a viable alternative which may account for faking more accurately than social desirability in job applicant settings. Finally, job desirability measures are more sensitive to faking than social desirability measures and should therefore be tested in applied settings.

**The Proposed Research: Purpose and Hypotheses**

The current study aimed to investigate and clarify various issues which, up to this point, have not been completely resolved in the faking literature. This section discusses each issue which the current study investigated, along with proposed hypotheses for testing each issue. The first two hypotheses addressed the prevalence of faking in applied
settings and the potential effects of faking on hiring decisions. The subsequent 4 hypotheses investigated the construct validity of job desirability measures and their potential for detecting job-specific applicant faking.

First, this study addressed whether or not faking does occur in applied settings. This was accomplished by addressing a major obstacle in faking research – the need for more applied samples of job applicants and incumbents. It is often difficult to obtain actual selection data in any area of I/O Psychology research, and even more difficult to assess whether some type of applicant faking occurred and affected hiring decisions or test validity. Therefore, only a limited amount of studies have investigated this issue in applied settings, and results have been contradictory, as discussed in the preceding review (e.g., Hough et al., 1990; Rosse et al., 1998). The only way in which researchers can eventually determine whether faking actually does occur in applied settings is to capture it in actual selection settings. This sentiment regarding the need for faking studies in applied settings was echoed by the chair of a symposium on applicant faking at the 17th Annual Meeting of the Society for Industrial and Organizational Psychology (SIOP) in Toronto. At the conclusion of the symposia presentations, she concluded with the remark that although many of the theoretical and methodological ideas proposed and tested successfully (mostly in laboratory settings) were promising, it was still necessary for these ideas to be applied to actual job applicants in order to ultimately assess their utility (Ryan, 2002). Thus, laboratory studies have provided much insight and testable hypotheses in regards to faking research, but in and of themselves, they cannot provide the answer to this particular question.
Hence, this study contributed to the existing faking research by assessing whether job applicants in actual selection settings fake on non-cognitive measures. As discussed earlier, many applied studies suggest that applicants fake in selection settings (Griffith et al., 2007; Birkeland et al, 2006; Weekley et al., 2003). One of the purposes of this study was to add to the extant literature by confirming these previous findings. However, the majority of these applied studies used traditional faking methods, such as social desirability scales, to measure faking. Few of these applied studies used alternative measures of faking, and even fewer have used job desirability measures. For these reasons, the proposed study makes an important contribution to the faking literature by using a job desirability scale to measure faking in an applicant setting, and comparing it to a traditional social desirability scale.

In order to confirm findings from previous studies, this study examined responses to personality and faking measures from a sample of applicants who applied for sales and managerial positions and then compared these responses to those collected from a separate sample of job incumbents employed in the same positions. It was expected that applicants would be more motivated than incumbents to fake on all selection measures in order to increase the likelihood of being hired for the job they are applying for. Therefore applicant scores on the personality and faking measures should be higher than incumbent scores in for both sales and manager positions. For the purposes of this study, the faking measure was a job desirability scale developed specifically for sales and management positions for the purposes of this study. A traditional social desirability scale was also included for comparison purposes.
Hypothesis 1a: For both sales and management jobs, applicants will score significantly higher than incumbents on all personality measures.

Hypothesis 1b: For both sales and management jobs, applicants will score significantly higher than incumbents on job desirability measures.

An important part of this investigation was the test of whether faking adversely affects actual hiring decisions to any substantial, and practical, extent. As previously discussed, validity coefficients for a whole sample of applicants may not decrease for various reasons, but hiring decisions can still be affected if a large percentage of fakers rise to the top of the applicant pool and are subsequently hired because of their high scores on a selection instrument such as a personality test (Mueller-Hanson et al., 2003; Rosse et al., 1998; Zickar et al., 1996). To assess whether this is the case, this study examined the extent to which applicants who fake on personality measures rise to the top of the personality score distribution and thus are over-represented at the high end of the applicant pool. Specifically, this study assessed whether the proportion of fakers at selection ratios of 50% or lower (more selective) is higher than the proportion of fakers in the entire sample of applicants. Fakers were operationalized as applicants who scored at the 95th percentile or higher on the job desirability scale that pertained to their job. In addition, this study assessed whether the percentage of fakers that would be selected using a top-down selection approach increases as the selection ratio becomes more selective. It is important to examine low selection ratios because the lower the selection ratio, the more selective a hiring organization can be. Thus actual organizations typically use low selection ratios in their hiring practices whenever possible (Cascio, 1998).
In effect, this portion of the current research makes a new contribution to the existing research on faking by examining whether fakers “rise to the top” of the applicant distribution using an alternative method to classify applicants as fakers – a job desirability scale. To date, no studies testing the “rise to the top effect” have used this type of faking measure. To further illustrate this point, in the Rosse et al. (1998), the authors identified fakers who rose to the top as those who scored highly on a measure of impression management (the BIDR-IM, Paulhus, 1991). Although this served as an excellent departure from the simple examination of validity coefficients, this method introduces the problems associated with using impression management scales which were described earlier. The other studies cited earlier which examined this “rise to the top” effect used different methods. Zickar et al. (1996) and Douglas et al. (1996) used computer simulations to model what the effects would be if applicants raised their scores on personality measures by various differing amounts. Therefore, in these studies, the personality measures themselves were used as the faking measures in these studies. And as for the Mueller-Hanson et al. (2003) study, faking was equated with being a participant in the incentive condition of a laboratory study, which implicitly encouraged respondents to fake. Therefore, none of the studies which investigated the “rise to the top” effect actually used any external measure of faking to classify respondents as fakers. The current builds upon this existing line of research by attempting to replicate the “rise to the top” effect while using a job desirability measure to identify fakers and assess whether they disproportionately increase their chances of being selected relative to non-fakers. It was predicted that at all selection ratios of 50% or lower, fakers would be overrepresented
in comparison to non-fakers. In addition, this overrepresentation was expected to increase as the size of the selection ratio decreased.

*Hypothesis 2a:* For both sales and manager applicants, at selection ratios of 50% or lower, the proportion of fakers will be disproportionately higher in the group of people selected than the proportion of fakers in the full sample. Fakers will be identified as applicants who scored at the 95th percentile or higher on the job desirability scale.

*Hypothesis 2b:* For both sales and manager applicants, as the selection ratio decreases, the proportion of fakers that will be hired will increase. Fakers will be identified applicants who scored in the 95th percentile or higher on the job desirability scale.

Job desirability is a response bias that, by definition, should be job specific (Kluger & Collela; Miller & Tristan, 2002). In other words, personality traits or experience qualifications which are relevant to one job may not be as relevant or desirable for a job which is very different, which has been shown to be the case when people are instructed to fake towards a specific job description (Mudgett, 2000; Raymark, 2004). Hence, an applicant for a particular job should only fake on job desirability items which are relevant to that job. Such a finding would lend important evidence of construct validity for the job desirability measure being used in this study, and on a broader scale, to the general construct of job desirability itself. This question was examined in this study by assessing whether sales applicants fake more on sales job desirability items than on management job desirability items, and vice-versa for manager applicants. It was posited that applicants
would fake more on job desirability items that were relevant to their job (sales or management) than on items that were not relevant to their job.

_Hypothesis 3a:_ Sales applicants’ mean score on the sales job desirability scale will be significantly higher than their mean score on the manager job desirability scale.

_Hypothesis 3b:_ Management applicants’ mean score on the manager job desirability scale will be significantly higher than their mean score on the sales job desirability scale.

Although this hypothesis provided a valuable within-subjects test of whether applicants fake more on job-specific items, it could be possible that applicants for other jobs fake to the same extent on the same items, indicating that faking may not be consistently job-specific. In order to rule out this possibility as well, a between-subjects test comparing sales and manager applicants scores on each of job desirability scales also was conducted. It was hypothesized that sales applicants would fake more than manager applicants on the sales job desirability scale, whereas those applying for managerial positions would outscore sales applicants on the manager job desirability scale.

_Hypothesis 4a:_ Sales applicants will score significantly higher than manager applicants on the sales job desirability scale.

_Hypothesis 4b:_ Manager applicants will score significantly higher than sales applicants on the manager job desirability scale.

Although job desirability is different from social desirability, the impression management component of social desirability does share a common characteristic with job desirability – it is a deliberate attempt to portray oneself in a favorable light (Paulhus,
Although job desirability aims at producing a portrait of a perfect applicant for a specific job, whereas social desirability involves a portrait of an all-around good person in general, they are both deliberate ways of faking on non-cognitive measures. One study in particular reported a low to moderate correlation between job desirability and the impression management factor of social desirability, but very little if any relationship between job desirability and the self-deception factor of social desirability, as this factor does not represent deliberate response distortion (Carroll et al., 2004). Moreover, Griffith et al. (2004) reported that neither impression management nor self-deception correlated significantly with faking (as measured by score differences between faking and honest conditions). The current study attempted to shed additional light on these findings by investigating in a field setting the relationships between faking and the two components of social desirability. It was hypothesized that the job desirability measure used in this study would correlate moderately with impression management, but not with self-deception. Confirmation of this finding would display further evidence of construct validity for job desirability and strengthen the notion that it captures deliberate deception rather than true personality variance.

**Hypothesis 5a:** In applicant data for both job types, there will be a low to moderate correlation between job desirability and the impression management subscale of social desirability.

**Hypothesis 5b:** In applicant data for both job types, there will be a negligible correlation between job desirability and the self-deception management subscale of social desirability.
As was discussed in the review, many job desirability measures have used a bogus item approach, where respondents are asked whether they have experience or familiarity with non-existent tasks, training, or equipment. This approach has proven to be fruitful as far as detecting applicant faking, since all answers are verifiable. However, one concern with bogus items is that applicants may truly believe that they have experience with these bogus tasks, training, or equipment because they are confusing them with ones that really do exist (Carroll et al., 2004). This was one concern expressed by the discussant at a faking symposium presented during the 2004 Society for Industrial & Organizational Psychology (SIOP) conference (McDaniel, 2004). For example, if the name of a bogus training program is similar enough to an actual training program that the applicant has undergone, the applicant may respond that they have this type of training when in fact they do not. Although this person would not be intentionally faking in this case, that item would label him or her as a faker simply due to an honest misunderstanding on the part of the applicant. This would be an example of the faking measure resulting in a false positive (Carroll et al., 2004).

In order to assess whether any misunderstandings such as these may have contributed to job desirability scores indicative of faking, incumbents’ job desirability scores were examined. If results indicated that applicants indeed faked their responses whereas incumbents did not, the mean job desirability score for applicants should be higher than that of incumbents, as is posited by Hypothesis 1. However, in order to show that the job desirability measure is not detecting false positives, the average score for incumbents should be extremely low, or close to zero. Assuming that job incumbents have no motivation to elevate their scores by faking, and if all honest responses to bogus items are
scored as a zero, then incumbents should have a minimal score (not significantly greater than zero) on a job desirability measure consisting of bogus items, such as the one which will be used in the current study. This finding would indicate that the job desirability measure results in a minimal amount of false positives. In comparison, applicants’ mean job desirability score should be much higher than this and would be significantly greater than zero. If the job desirability measure used in this study did result in a minimal amount of false positives, it would provide additional construct validity in the sense that it is measuring actual applicant deception, rather than unintended errors on the part of applicants. Therefore it was expected that for incumbents, scores on the job desirability measure would approach zero whereas applicants’ mean scores would be significantly greater than zero.

**Hypothesis 6a:** Incumbent mean scores on the job desirability scales will not be significantly greater than zero.

**Hypothesis 6b:** Applicant mean scores on the job desirability scales will be significantly greater than zero.

It is quite plausible that job experience in a similar line of work could help an applicant distort their responses in a more job-specific fashion than someone who knows little about the target job. More importantly, it is possible that a person with highly relevant experience may also fake in a more subtle manner, making it difficult to detect distortion because he or she does not need to fake on every item, but only on job-relevant items. This could be due to job-related knowledge gained over several years of performing a specific job and multiple prior experiences applying for such jobs, in which
instances a person would encounter several types of selection instruments which are used to select applicants for that type of job.

Therefore, it is possible that applicants with greater amounts of experience in sales or management are more likely to distort their responses in a job-desirable manner. This finding would give a further indication that job desirability is highly job specific, thus lending further evidence of construct validity for job desirability.

Due to the novelty of job desirability measures in faking research, no studies to date have investigated whether job-specific faking behavior can in any way be predicted by years of job experience. Hence, this analysis also contributed to the existing faking research. It was predicted that job desirability scores would correlate with years of job experience in the specific job that applicants were applying for.

Hypothesis 7: For applicants, years of job experience in a similar position will be significantly and positively correlated with job desirability scores.
III. METHOD

Participants

A total of 1,780 job applicants and incumbents participated in the study. However, only 958 of these were applying for, or employed in, sales or manager positions. The remaining 822 participants were individuals applying for other types of positions, which were not within the scope of this study. Therefore data from these individuals were not included. It should be noted that this study specifically excluded any individuals applying for, or currently employed in, sales manager positions. This was done in order to minimize any possible confounding effects due to positions that combined both sales and manager requirements.

For the purposes of this study, the applicant sample consisted of 632 sales job applicants and 112 management job applicants from various cities throughout the U.S. The incumbent sample consisted of 130 individuals currently employed in sales positions and 84 currently employed in management positions. Data from these participants was gathered through a private management consulting firm located in the Midwestern U.S., which carries out web-based assessments for multiple client organizations for the purposes of selection and employee development. A subset of the sales incumbent sample was provided by a second consulting firm located in the Eastern U.S. which also carries out similar web-based assessments for employee selection and development. Sales incumbents from this subset underwent a similar web-based assessment for the same purposes. However, due to time constraints, this sample of participants only completed
one of the measures included in this study. All applicants who participated were being assessed as part of the hiring process for a position (sales or manager) they were seeking with an organization. However, all incumbents who participated were being assessed by their current employer only for career development and training purposes. Their results were not used for promotional or selection/de-selection purposes.

Measures

Demographics. A measure was administered to all participants that asked them to voluntarily provide their age, gender, race, and years of overall experience working in a job similar to what they were currently applying for or employed in.

Personality Measures. Goldberg’s International Personality Item Pool (IPIP) scales were chosen as the principal measure of personality in this study because of their high correlations with two other well-known Big Five personality measures, the NEO-PI-R (McCrae & Costa, 1992) and the Hogan Personality Inventory (Hogan, 1992). These high correlations were obtained in a study which used a sample of 501 subjects who completed several published personality measures in addition to the IPIP scales (Goldberg, 1999).

In this study, two of the IPIP Big Five scales were used – the Conscientiousness and Extraversion scales. These scales are both five-point graphic rating scales with a response format as follows: 1 = Very Inaccurate, 2 = Moderately Inaccurate, 3 = Neither Inaccurate or Accurate, 4 = Moderately Accurate, and 5 = Very Accurate. In order to avoid adding excessive testing time to the consulting firm’s assessment, the 10-item versions of these two IPIP scales were used, as opposed to the longer 20-item versions.
The 10-item IPIP Conscientiousness scale has an alpha of .81, compared with an alpha of .71 for the NEO-PI-R Conscientiousness scale, and correlates .79 (.92 when corrected for attenuation due to scale unreliability) with the NEO-PI-R Conscientiousness scale. The 10-item IPIP Extraversion scale has an alpha of .86, compared with an alpha of .76 for the NEO-PI-R Extraversion scale, and correlates .77 (.88 corrected) with the NEO-PI-R Extraversion scale. Therefore, it is apparent that the shorter IPIP scale versions still retain satisfactory psychometric quality.

In addition, two IPIP scales corresponding with the Hogan Personality Inventory (HPI) occupational scales were included in the study. The first occupational scale - the IPIP Gregariousness scale - correlates .73 (.84 corrected) with the HPI Sales Potential scale and has an alpha of .86. The second occupational scale used was the IPIP Competence scale, which correlates .62 (.78 corrected) with the HPI Managerial Potential scale and has an alpha of .80. Both of these scales consist of 10 items, and they have the same 5-point graphic scale response format as the IPIP Conscientiousness and Extraversion scales described above.

Social Desirability Measure. To measure social desirability, the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984) was used. This measure consists of two subscales, Impression Management and Self-Deception, which are each 20 items in length. The Impression Management subscale has an alpha ranging from .75 to .86 and the Self-Deception subscale has an alpha ranging from .68 to .80. The response format for all items is a seven-point graphic rating scale ranging from 1 = Not True to 7 = Very True, with representative anchors between these two responses. However, the scale
is scored dichotomously. Only extreme responses of 1, or 2, count as socially desirable responses.

*Job Desirability Measure.* To measure job desirability, two scales were developed for this study. One scale targeted sales jobs and the other targeted management jobs. Each scale consisted of 12 bogus items that asked participants whether they had experience, or were familiar with, various terms or procedures that were superficially relevant to the particular job that they were applying for, but were in fact non-existent. An example of such an item is, “I have read Tony Robinet’s book entitled *Sales for the Third Millennium.*” Such a book does not exist; therefore it is verifiably impossible to have read it. All items had three response options: "True", "False", and "Not Sure.” This scale also was scored dichotomously. Only items endorsed as “True” were counted as faked responses. Items endorsed as “False” and “Not Sure” were not scored as faked responses. A very similar job desirability measure constructed for a previous laboratory study on the same topic (Miller & Tristan, 2002) displayed adequate reliability in an honest-instruction condition (alpha = .70) and highly satisfactory reliability in a faking-instruction condition (alpha = .98).

In order to increase the realism of these items to applicants and help conceal their true purpose (to detect faking), these bogus items were combined with items that asked them about their experience with actual terms or procedures that did exist and were either related to sales or management jobs. Eight of these “filler” items were added to both the sales and manager job desirability scales. Therefore, the total number of items comprising each scale was actually 20 items, but only 12 of the items actually comprised each scale and subsequently were included in the analyses.
As noted previously in the Hypotheses section, faking in this study was operationalized as having scored at the 95th percentile or higher on the job desirability scale. As such, sales applicants who scored in the 95th percentile or higher on the sales job desirability scale were identified as fakers in the sales group. The same procedure was carried out for the manager applicant group using the manager job desirability scale scores.

**Procedure**

**Description of Assessment Process.** Participants completed all of the measures included in this study concurrently with the consulting firm’s web-based assessment. The assessment procedure of the consulting firm for selection projects is such that job applicants who wish to apply for a given job with a client company (of the consulting firm) must undergo the consulting firm’s assessment on-line over the internet. All applicants are assigned log-in identification numbers by the organization whom they are applying with and instructed to take the assessment on their own time from any given location. All applicants then complete the assessment. Once the entire assessment is completed, the consulting firm makes a recommendation to the client organization, which in turn, uses this recommendation as part of their own overall selection process in order to make a hiring decision regarding each applicant. Each firm’s instructions and assessment were administered in full prior to administration of the instructions and measures that were part of this study.

As noted earlier, a second consulting firm provided data for part of the incumbent sample. These data were only for sales incumbents, however. Both firm’s procedures for development assessments are the same as for their selection assessments, except for the
fact that those being assessed are employees working for client companies. The purpose of these assessments is not for selection, but for organizations to assess their current talent for a variety of purposes, such as career development, needs assessment, and succession planning. It should be noted that the sales incumbent participants from this subset only completed the Sales Job Desirability measure. Due to time constraints, they were not administered any of the personality or social desirability scales, nor were they administered the Manager Job Desirability scale.

*Instructions.* When participants logged in to take the assessment, they were presented with either one of the consulting firm’s standard on-line assessment instructions, which are shown in Appendix D, followed by the assessment of the firm. The assessment processes used for this study did not differ from either consulting firm’s standard administration process in any manner aside from the inclusion of the measures pertaining to this study, which were added in a final section at the end of each firm’s standard assessment. All testing instructions pertaining to this study, also shown in Appendix D, were the same regardless of the consulting firm.

*Debriefing and Informed Consent.* Originally this study was planned as a within-subject design quasi-experiment, where participants would complete all measures as applicants (Time 1) and would then volunteer to complete the same measures again as employees (Time 2) after being hired by an organization. As such, a debriefing statement was to be administered to participants after they had completed the second administration of the assessment. However, near the end of the Time 1 data collection phase, it was learned that a Time 2 data collection phase would not be possible. Hence, the study design was modified into a between-subjects design comparing applicants to incumbents.
As a result, the participants in the applicant group did not receive a debriefing statement after they completed the measures in this study. In order to keep consistency across applicant and incumbent groups, a debriefing statement was not provided to participants in the incumbent group.

Additionally, it should be noted that due to the nature of this applied study, which compared effects of an applicant versus an incumbent setting on test scores, it was crucial that applicants believed that all the measures that they were completing were being used for actual hiring recommendations. Therefore, informed consent was not collected for applicants in this study. In order to maintain test setting consistency across applicant and incumbent samples, it was not collected for incumbents either. Due to the fact that the measures administered in this study were very similar in nature to those being administered by the firm, and the fact that the addition of measures in this study only added approximately 15 minutes of assessment time without inducing any additional stress or harm to participants, the omission of informed consent at this initial stage did not violate any ethical standards for psychological testing.
IV. RESULTS

Demographics

The average age of participants was 34.52 years (SD = 10.24), with a range of 19 to 66 years. The sample consisted of 570 (59.5%) male participants and 360 (37.6%) female participants. Seventy-one percent of participants were Caucasian, 9.7% African American, 7% Asian, and 5.4% Hispanic. Over half (52.6%) of the sample possessed 5 or less years’ experience in a job similar to what they were currently applying for or employed in (either sales or management) and 21% possessed 10 or more years of experience. Eighty percent of participants were either applying for, or currently employed in, sales jobs, whilst 20% were applying for, or employed in, management jobs. Individuals applying for alternative positions, such as sales managers, professional and technical, customer service, and administrative, also participated in the study. As noted earlier, however, their data were not included as part of the study and are not included in the aforementioned demographic statistics. Table 1a lists demographic data only for participants who were included in the study whereas Table 1b lists demographic data for all individuals who participated, regardless of their inclusion in the study.

Scale Reliabilities

Scale inter-correlations and reliabilities for all scales are presented in Tables 2a through 2g for the following sample subsets: full sample, sales only, managers only, sales applicants, sales incumbents, manager applicants, and manager incumbents. Reliability
coefficients for all measures were acceptable, ranging from .75 to .83 in the full sample and ranging from .63 to .89 across these other sample subsets.

Tests of Hypotheses

Hypothesis 1. Hypothesis 1a predicted that, for both sales and manager jobs, applicants would score significantly higher than incumbents on all personality scales. The four personality scales were Conscientiousness, Extraversion, Competence, and Gregariousness. In order to test this hypothesis, independent measures $t$-tests were conducted to compare applicant versus incumbent mean scores on the four personality scales for both job types. As shown in Table 3a, sales applicants scored significantly higher than sales incumbents on all four personality scales. However, manager applicants scored significantly higher than manager incumbents only on two of the four personality scales - Conscientiousness and Competence (see Table 3b). Therefore Hypothesis 1a was partially supported.

An examination of scatter plot graphs revealed the scores on all personality scales appeared negatively skewed. Tests for skewness and kurtosis were conducted in order to assess the degree to which these scores were non-normally distributed. Results from skewness tests confirmed that all personality scale scores exhibited a degree of non-symmetrical distribution which indicated a departure from normality based on the fact that skewness values for all scales were more than twice the magnitude of their standard errors (Tabachnick & Fidell, 1996). In addition, tests for kurtosis were conducted to assess whether the clustering observed in personality scores suggested non-normality. Kurtosis values also were more than twice the magnitude of their standard errors for all scales, particularly for Conscientiousness and Competence scores. This finding provided
further evidence of non-normality in the personality scores. Hence based on tests for skewness and kurtosis, scores on all four personality scales were deemed to be non-normally distributed, both in the full sample and in the sales and manager samples.

In order to avoid violations of normality, nonparametric tests were also conducted to test Hypothesis 1a (Tabachnick & Fidell, 1996). A series of Mann-Whitney $U$-tests were thus conducted to test whether applicant scores on each of the scales were significantly higher than incumbent scores for both sales and manager jobs. As shown in Table 3a, results from the Mann-Whitney $U$-tests differed only slightly than results from the $t$-tests. Although sales applicants outscored sales incumbents on all scales, these differences were significant only for Extraversion and Gregariousness. Conversely, in the manager group applicants outscored incumbents on all scales but these differences were only significant for Conscientiousness and Competence (see Table 3b). Therefore, Hypothesis 1a also was partially supported using a nonparametric version of the independent measures $t$-test. Results were highly similar to those obtained with a parametric test and differed only in terms of significance levels for two of the scales in the manager job group.

Hypothesis 1b predicted that, for both sales and manager jobs, applicants would score significantly higher than incumbents on the job desirability scale that was developed for their job group (sales or manager). In order to test this hypothesis, independent measures $t$-tests were conducted to compare applicant versus incumbent mean scores on the job desirability scales for both job groups. Table 3c illustrates that in the sales group, applicants did not score significantly higher than incumbents on the sales job desirability scale. Although applicants had higher mean scores, this difference only approached
statistical significance \((t = 1.77, p = .08)\). For managers, however, this hypothesis was supported in that applicants scored significantly higher than incumbents \((t = 3.77, p < .001)\) on the manager job desirability scale (see Table 3d). Therefore Hypothesis 1b was partially supported.

As with Hypothesis 1a, skewness and kurtosis tests revealed that job desirability scores were also non-normally distributed. This was evidenced by high positive skewness values for the sales and manager job desirability scales in the full sample and in both the sales and manager sample subsets. These positive skew values were all well above twice the magnitude of their standard errors. Kurtosis values also indicated some degree of non-normality in the job desirability scores. Therefore, nonparametric tests were also conducted to test Hypothesis 1b.

Mann-Whitney \(U\)-tests confirmed findings from the independent measures \(t\)-tests (see Tables 3c and 3d). Sales applicants did not score significantly higher than sales incumbents on the sales job desirability scale, although they did score higher and this difference approached statistical significance \((Z = -1.86, p = .06)\). As for managers, applicants did score significantly higher than incumbents \((Z = -3.86, p < .001)\) on the manager job desirability scale, providing partial support for Hypothesis 1b using nonparametric tests.

Tables 11a and 11b provide descriptive data each item comprising the sales and manager job desirability scales, respectively. In addition, this table provides standardized mean differences in item means between the sales and manager applicant groups.

**Hypothesis 2.** Hypothesis 2a predicted that, for both sales and manager applicant groups, at selection ratios of 50% or lower, the proportion of fakers selected would be...
disproportionately higher than the proportion of fakers in the full sample. In order to test Hypothesis 2a, the first step was to identify the fakers in the sales applicant group and the fakers in the manager applicant group. Sales applicants were classified as fakers if their sales job desirability score was at, or above, the 95th percentile of job desirability scores within the sales applicant sample. The same procedure was conducted for manager applicants using the manager job desirability score in order to identify fakers in this group. This procedure resulted in classifying all sales applicants with scores of seven or higher as a sales faker. In other words, any sales applicant who responded “yes” to seven or more of the 12 sales bogus items was classified as a faker. For managers, those with scores of six or higher (responding “yes” to six or more of the 12 manager bogus items) were classified as manager fakers. Previous faking studies have used the 95th percentile as a cut-off using social desirability scales in order to classify respondents or applicants as fakers (Ellingson et al., 2001; Rosse et al., 1998), therefore the same cut-off was used for the job desirability measures in this study.

Next, the percentage of fakers who would be hired (based on a top-down rank ordering of personality scores) from each job applicant group were calculated. These percentages were calculated for each of the four personality scales using the following selection ratios: .50, .25, .10, and .05. These resulting percentages of fakers who would be hired were then compared to the percentage of fakers in the whole sample of applicants in order to assess whether each percentage was higher than the percentage of fakers in the full sample. As shown in Tables 4a and 4b, results indicated that at any selection ratio of .50 or lower, the proportion of fakers who would be hired was higher than the proportion
of fakers in the full sample. These results were found for both sales and manager jobs, thus Hypothesis 2a was fully supported.

Hypothesis 2b predicted that for both sales and manager applicant groups, as the selection ratio decreased (thus becoming increasingly selective), the proportion of fakers hired would increase. This hypothesis was supported, as shown in Tables 4a and 4b. In both job groups, the percentage of fakers that would be hired increased as the selection ratio decreased. The proportion of fakers consistently increased in a linear fashion for all personality scales from selection ratios of .50 through .10 although they held constant from selection ratios of .10 to .05.

Hypothesis 3. Hypothesis 3 provided a within-subjects test of whether applicant faking on the job desirability measures would display a job-specific pattern by assessing whether each applicant group (sales or manager) would score significantly higher on the job desirability score that pertained to their job group. In order to test Hypothesis 3a, a dependent-measures t-test was used to assess whether sales applicant mean scores on the sales job desirability measure were significantly higher than their mean scores on the management job desirability measure. Results, displayed in Table 3, indicated that sales applicants scored significantly higher ($t = 11.78, p < .001$) on the sales job desirability scale than they did on the manager job desirability scale. Using a Wilcoxon’s Signed Ranks test, a nonparametric version of a dependent-measures t-test, these results were replicated ($Z = -10.75, p < .001$). Thus, Hypothesis 3a was fully supported both parametric and nonparametric tests.

Similarly, to test Hypothesis 3b, a dependent-measures t-test was used to assess whether manager job applicants’ mean score on the management job desirability measure
was significantly higher than their mean score on the sales job desirability measure. Results indicated that although managers scored higher on the manager scale than on the sales scale, this difference was not statistically significant ($t = 0.97, p > .05$). As similar results were found using a Wilcoxon’s Signed Ranks test ($Z = -0.94, p > .05$), Hypothesis 3b was not supported. Consequently, Hypothesis 3 received only partial support.

Hypothesis 4. In contrast to the previous hypothesis, Hypothesis 4 provided a between-subjects test of whether faking was job-specific on the job desirability measures. Hypothesis 4a posited that sales applicants would score significantly higher than their manager counterparts on the sales job desirability scale. To test this hypothesis, an independent-measures $t$-test was conducted to assess whether sales applicants’ mean score on the sales job desirability measure was significantly higher than that of management applicants. Results did not support this hypothesis, however. Sales applicants scored higher than manager applicants on the sales job desirability scale, however, the difference only approached statistical significance ($p = .07$). Nonetheless, when a Mann-Whitney $U$-test was utilized to test this hypothesis, the difference was statistically significant ($Z = -2.33, p < .05$), with sales applicants scoring significantly higher on sales job desirability than manager applicants.

Hypothesis 4b was complimentary to Hypothesis 4a in that it predicted that manager applicants would outscore sales applicants on the manager job desirability scale. In order to test this, an independent-measures $t$-test was used to assess whether manager applicants’ mean score on the manager job desirability scale was significantly higher than that of sales applicants. It was found that manager applicants scored significantly higher ($t = -2.88, p < .01$) than sales applicants on the manager job desirability scale. Moreover,
these results also were found using a Mann-Whitney U-test \((Z = -3.34, p < .01)\), offering support for Hypothesis 4b. As a result, Hypothesis 4 was fully supported using nonparametric tests but only was partially supported using parametric tests. Full results are provided in Table 6.

Together, Hypotheses 3 and 4 provided separate within- and between-subject tests of whether faking on the job desirability measures was job specific. As discussed, results from the separate \(t\)-tests suggested that this was the case, which indicated the presence of a job group (sales or manager) by job desirability scale version (sales or manager version) interaction. In order to confirm the presence of such an interaction, it was decided to conduct a follow-up analysis to Hypotheses 3 and 4 using an Analysis of Variance (ANOVA), which provided a more statistically powerful and parsimonious method with which to test the combined effects of job type and scale version on job desirability scores. Thus, a 2 (Job Desirability Scale Version) x 2 (Job Group) mixed-design ANOVA was conducted. The two independent variables were Scale Version and Job Group, where Scale Version was the repeated measures factor and Job Group was the between-subjects factor. Scores on the two job desirability scales were used as the dependent variable.

It was expected that a significant Scale Version x Job Group interaction would be observed whereby scores on the two versions of the job desirability scales would significantly differ by job group. Prior to conducting this analysis, a test for homogeneity of variance was conducted to assess whether the error variance of the job desirability scores was equal across sales and manager groups. A Levene’s Test of Equality of Error Variances indicated that the data did not violate any assumptions of homogeneity for either the sales \([F (1,741) = .22, p = .64]\) or manager groups \([F (1,741) = 1.36, p = .24]\). It
was not necessary to conduct tests for sphericity due to the fact that there were only two levels of the repeated measures variable (sales and manager version of the job desirability scale).

Results of this analysis revealed that the main effect for job group was not significant \([F (1,741) = .15, p > .05, \text{Eta-squared} = .00]\). Thus there was no significant difference between sales applicants \((M = 1.94)\) and manager applicants \((M = 2.02)\) on overall job desirability scores. A significant main effect for Scale Version was obtained, \([F (1,741) = 13.82, p < .001]\), though this was a weak effect \((\text{Eta-squared} = .02)\). Applicants overall scored higher on the sales job desirability scale \((M = 2.15)\) than they did on the manager scale \((M = 1.80)\). However, a significant Scale Version x Job Group was also observed, \([F (1,741) = 29.63, p < .001, \text{Eta-squared} = .04]\). Examination of the cell means indicated that applicants’ scores on the two versions of the job desirability scale differed based upon the type of job they were applying for. Specifically, sales applicants \((M = 2.37)\) scored higher on the sales job desirability scale than they did on the manager scale \((M = 1.51)\). However, the reverse pattern was observed for manager applicants, who scored higher on the manager job desirability scale \((M = 2.10)\) than they did on the sales scale \((M = 1.94)\). A graphical representation of this interaction is shown in Figure 1. The results of this follow-up analysis thus support the findings from Hypotheses 3 and 4 by providing evidence that responses on the job desirability scales differed according to the type of job that applicants were seeking. Results for this follow-up analysis are presented in Table 7.

**Hypothesis 5.** Hypothesis 5 addressed the degree of association between job desirability and the two types of social desirability – impression management and self-
Hypothesis 5a posited that job desirability would be significantly correlated with impression management in each applicant job group. In order to test Hypothesis 5a, Pearson’s $r$ correlation coefficients were calculated in order to assess whether the correlation between job desirability and impression management for applicants in both job groups was statistically different from zero and of moderate strength. For the sales applicant group, the correlation between the sales job desirability scale and the impression management scale was statistically significant and positive ($r = .12, p < .01$) whilst for the manager applicant group, the manager job desirability scale also was significantly and positively related to the impression management scale ($r = .30, p < .01$). Thus, Hypothesis 5a was supported.

Hypothesis 5b predicted that job desirability would not significantly correlate with the self-deception scale. Pearson’s $r$ correlation coefficients again were conducted to test this hypothesis. Contrary to expectation, in the sales applicant group, sales job desirability significantly and positively correlated with self-deception ($r = .21, p < .001$). Likewise, in the manager applicant group, the manager job desirability scale also was significantly and positively correlated with self-deception ($r = .35, p < .001$). As these results failed to support Hypothesis 5b, only partial support was found for Hypothesis 5.

Hypotheses 5a and 5b also were tested using Spearman’s rho correlation coefficients due to the previously discussed non-normal distribution of the job desirability scores. These results closely mirrored those found using Pearson’s correlations, thus lending the same partial support for Hypothesis 5. Full results using both types of correlational analyses are shown in Tables 8a and 8b.
Hypothesis 6. Hypothesis 6 investigated whether job incumbents engaged in any discernable amount of faking in comparison to job applicants. As such, Hypothesis 6a predicted that incumbents would have mean scores on the job desirability measures that were not significantly different than zero whereas Hypothesis 6b predicted that applicant scores on the same measures would be significantly higher than zero. In order to test these hypotheses, for both job groups, a one-sample \( t \)-test was calculated using incumbents’ mean job desirability score to assess whether it is significantly different from zero. The same procedure was then conducted for applicants from both job groups. A statistically insignificant finding would indicate that the job desirability scale resulted in a minimal amount of false positives, or individuals whom were falsely identified as fakers.

Results did not support Hypothesis 6a. Sales incumbents scored significantly higher than zero \((t = 10.60, p < .001)\) on the sales job desirability scale, as did manager incumbents on the manager job desirability scale \((t = 6.73, p < .001)\). This suggested that incumbents from both job groups actively engaged in at least some level of faking.

As expected, however, Hypothesis 6b was supported. Sales applicants scored significantly higher than zero \((t = 26.09, p < .001)\) on the sales job desirability scale and manager applicants scored significantly higher than zero \((t = 10.10, p < .001)\) on the manager job desirability scale. These analyses were then replicated using a one-sample Kolmogorov-Smirnov test, which is a nonparametric version of the one-sample \( t \)-test. Results from these nonparametric tests revealed the same findings; incumbents and applicants from both job groups all had mean job desirability scores that were significantly higher than zero. Hence Hypothesis 6 received only partial support.
regardless of using parametric versus nonparametric test use. Full results are displayed in Tables 9a and 9b.

Hypothesis 7. Hypothesis 7 predicted that job experience would be positively correlated with the extent of faking observed. In order to test Hypothesis 7, for both applicant job groups, Pearson’s correlation coefficients were calculated in order to assess whether job experience in a similar job was significantly and positively correlated with job desirability scores. Results provided partial support for Hypothesis 7. In the sales applicant sample, the sales job desirability scale correlated significantly and positively ($r = .26, p < .001$) with job experience. However, this was not the case in the manager applicant sample, where only a weak and non-significant positive correlation was found between the manager job desirability scale and job experience ($r = .07, p > .05$).

Highly similar results were found using Spearman’s rho correlations. Thus, only partial support was found for Hypothesis 7 using both parametric and non-parametric analysis methods. Full results are shown in Table 10.
V. DISCUSSION

Although the current study confirmed some of the previous findings in the literature on applicant faking, it made various unique contributions as well. The findings from Hypothesis 1a replicated findings from multiple studies in which applicants have scored significantly higher on personality measures than incumbents using between-subjects designs in an applied setting (Birkeland et al., 2006; Robie et al., 2001; Weekley et al., 2003). Sales applicants scored significantly higher than sales applicants on the four personality measures (only two scales when non-parametric tests were used). Manager applicants also tended to score higher than manager incumbents, although these differences were significant for only two of the four scales, regardless of significance test type. Thus, differences in personality scores reflected a degree of faking on the part of applicants, although the magnitude of this bias varied depending on the job type and the personality scale in question.

Upon further review, the observed pattern of differences on the personality measures indicated that applicants increased their scores to a greater extent on the scales which were most relevant to the job they were applying for. To illustrate, the effect sizes for the sales group in Hypothesis 1 (see Table 3a) were much higher for the extraversion and gregariousness scales, which are stereotypically very relevant to sales jobs, indicating that sales applicants selectively faked more on these constructs. This finding corroborates Birkeland et al.’s (2006) recent meta-analysis findings, which reported that sales applicants inflate their scores more on extroversion compared to other Big Five
constructs. The authors attributed this finding to applicants’ perception that extroversion is related to success in sales. Conversely, the current study found that manager applicants elevated their scores noticeably more on the other two scales: conscientiousness and competence. This resulted in significantly higher scores and effect sizes (.54 for conscientiousness and .40 for competence, versus .13 for extraversion and .22 for gregariousness), which are shown in Table 3b. This is interesting, given that Birkeland et al. (2006) also hypothesized that managers would inflate their scores more on conscientiousness relative to other constructs but their results failed to support this. This study’s findings thus mirror those from earlier research, such as the previously cited study by Burkrant (2001), who found that applicants faked more on items that were relevant to the job they were applying for.

These results provide additional evidence that applicants fake in a manner that is job-specific. This growing body of research should be considered when interpreting recent conclusions made by authors such as Ellingson et al. (2007), which state that applicant faking is a non-significant issue. This particular study, which isolated the average effect due to faking from that of time and feedback, did not examine job type at all. Thus, mean effect sizes in applicant-incumbent score differences were averaged across job types. In addition, the estimated mean effect size due to faking was averaged across 20 different CPI personality scales. Upon further investigation, the effect sizes for these scales ranged widely (from less than 0.01 to as much as 0.40 standard deviations) depending on the construct measured by the scale. Given previously mentioned findings about job-specific faking patterns of applicants (e.g., Birkeland et al., 2006), it is quite
possible that much larger effect sizes would have been reported by the authors had they 1) examined effect sizes for each scale individually and 2) analyzed the data by job type.

One of the major unique contributions of this study was the use of an alternative faking measure in an applied setting. The sales and manager job desirability scales developed for this study expanded upon the type of bogus items that have been used in previous research. First, no previous study had used bogus item measures designed specifically for sales and manager positions and administered them in an applied setting. Of the bogus item-based measures previously used in applied studies, such as those created by Pannone (1984) and Anderson et al. (1984), all have been targeted towards lower-level positions, such as electrician and clerical/administrative, respectively. In addition, the two scales in this study were each comprised of 12 bogus items; in comparison, Pannone’s (1984) measure consisted of just one item. This likely contributed to the higher reliability values observed for the two scales, ranging from .73 to .77 across samples. The only exception was for the manager incumbent sample, in which values ranged from .63 to .65.

Based on the body of literature discussed earlier, it was expected that applicants from each job group would score significantly higher than incumbents on their specific job desirability scale, which would have supported Hypothesis 1b. Indeed this was the case for the manager group, in which applicants significantly outscored incumbents by a difference of over one-half standard deviations. Surprisingly, this was not the case for the sales group. Sales applicants scored higher than sales incumbents, but this difference was not statistically significant. One would initially conclude from this observation that sales applicants simply did not fake in this instance. Upon closer inspection this conclusion is
premature. An examination of the mean scores for sales applicants and sales incumbents indicates that the lack of significant findings was due more to the surprisingly high mean score of sales incumbents (1.98) rather than to a low mean score of sales applicants (2.37). In other words, sales incumbents had unexpectedly high scores on the faking measure, which resulted in a small mean applicant-incumbent difference. Sales applicants still engaged in faking, but sales incumbents faked as well, thus precluding any significant difference in faking scores. In contrast, manager incumbents had a mean score of only 1.06 on the manager job desirability scale, compared to their sales applicant counterparts, who had a mean score of 2.06. As shown in Tables 3c and 3d, this resulted in a much larger applicant-incumbent effect size for the manager job (0.57) than for the sales job (0.18).

This finding has important implications for applied faking and selection research. The first implication is that sales may be a unique position that does not lend itself easily to applicant-incumbent comparisons due the nature of the job. As a result, individuals employed in sales positions may have a greater tendency to fake on personality measures in comparison to individuals employed in alternative positions that have no sales function. Sales roles inherently require a high degree of impression management behavior and skills, and success in many types of sales roles is highly dependent on making and maintaining a favorable impression. It may be that what some consider faking, lying, or dishonesty may simply be seen as putting the best possible foot forward to a salesperson. Research would suggest that being high in social desirability may in fact be related to strong sales performance. For example, Ruch and Ruch (1967) found that correcting the MMPI according to the K scale (internal validity scale) actually decreased validity in a
sample of sales managers. Moreover, previous research has found that the same individual will present him/herself differently when applying for a sales job versus non-sales jobs (Velicer & Weiner, 1975; Wesman, 1952). Thus faking research should examine the unique response tendencies of sales applicants and further investigate how faking affects test validity in this applicant population.

The second implication is that designs in applied faking research may rely too heavily upon the assumption that incumbents do not fake on non-cognitive measures. An overwhelming majority of the existing field studies on faking have used a between-subjects design that compares applicants to incumbents, such as was used in this study. Significant differences, or lack of, in mean scores are then interpreted as evidence that applicants fake (or do not fake) in selection settings. This design is based on the assumption that incumbents do not engage in faking simply because they do not have the same motivation to elevate their scores as applicants do. Although this is probably often the case, the findings of this study raise some important concerns about this assumption. In this study, incumbents from both groups engaged in faking, as indicated by results from the one-sample t-tests conducted for Hypothesis 6. These results indicated that incumbents had scores that were significantly higher than zero on the job desirability scales, despite the fact that they lacked the same motivation to fake as the applicants (improving their chances at being selected for a job they were applying for). It should be noted that incumbents in this study were being assessed for developmental purposes. Although they were not being considered for promotional or internal selection purposes, it is possible that there was nevertheless some level of motivation to fake in order to produce a desirable profile that was job-specific. To the extent that any such motivation
was present in the incumbents’ response set, this would have resulted in elevated job desirability scores for this group.

Hypothesis 2 replicated previous findings from applied studies that demonstrate how individuals who fake can rise to the top of an applicant pool, thereby greatly increasing their chances of being selected when a company uses a top-down selection strategy (Griffith et al., 2000; Rosse et al., 1998). At selection ratios of .50 or less, the proportion of fakers that would be selected was consistently higher than the proportion of fakers in the total sample. More importantly, the proportion of fakers who would be selected linearly increased as the selection ratio became more selective. This finding provides further evidence that faking can create powerful rank-order changes in applicant pools which increase the probability that an organization will hire an applicant who otherwise would not have been selected. In light of this, one can call into question the conclusions made by Ellingson et al. (2007) and Hogan et al. (2007) which purport that applicant faking is minimal and has inconsequential effects on hiring decisions. Neither of these studies examined the effects of faking at the top score percentile ranges - their results were based on the full distribution of scores. However, the results from this study and previous research show that applicant faking tends to dramatically increase at the higher percentiles of a personality score distribution. If Ellingson et al. (2007) had analyzed applicant-incumbent differences only for those who scored in the top 5 to 25 percent they may have reported much higher effect sizes even after accounting for time and feedback effects. This issue has even stronger implications for the Hogan et al. (2007) study findings because the study examined data only for applicants who were rejected for a position during their first job application because their personality scores
were too low. As a result of this sampling method, the applicants who were most likely to fake were inherently excluded from the study. Thus, it would be erroneous to conclude that applicants do not fake based on these findings, especially given previously established research confirming that there exist individual differences in faking (e.g., McFarland & Ryan, 2000). Hogan et al.’s (2007) findings likely were based on a limited range of applicants who had low ability or propensity to fake because the study systematically removed individuals who were more likely to fake (those who were successful in their first application attempt). For example, applicants with higher ability or propensity to fake, such as those lower in conscientiousness or locus of control (Griffith et al., 2004), would have been systematically removed in Hogan et al.’s (2007) study, leading to the conclusion that faking did not occur.

In addition to the fact that studies such as Ellingson et al. (2007) do not account for rank-order changes at the top of the applicant score distribution, they fail to examine the role of job type in faking. As noted earlier, the results of the current study showed that applicants inflated their scores most on personality factors that were most relevant to the job they were applying for (e.g., sales applicants inflated their scores more on extroversion than conscientiousness and vice-versa for manager applicants). As such, it is likely that much of the conflicting literature regarding the prevalence of applicant faking is due to the lack of attention given to job type. If applied faking studies examined applicant-incumbent or selection-development setting differences and compared them for different job types (rather than averaging across all job types), significant faking effects likely would be observed. The current study and studies such as Birkeland et al. (2006) are examples that support this argument. Future research should more rigorously
investigate the role of job type on whether applicants fake, the extent to which they fake, and identify which personality factors they are most likely to fake on given the type of job they are applying for. For example, based on this study’s findings and many previous studies (Birkeland et al., 2006; Ruch & Ruch, 1967; Wesman, 1952) it is evident that sales positions are fairly unique in terms of what personality factors are most likely to be emphasized or inflated during the application process.

The current study also provided a contribution to existing research by replicating the applicant “rise to the top” effect (the finding that applicants who fake increase their chances of being hired in a top-down selection system based on non-cognitive test scores) through the use of an external measure of faking. As discussed earlier, the studies by Rosse et al. (1996) and Zickar et al. (1996) relied on computer simulations of personality scores while Mueller-Hanson et al. (2003) classified participants as fakers simply based upon their assignment to an incentive condition in the study. Griffith et al. (2000) documented this effect simply by examining rank-order changes in the personality scores themselves. The fact that this effect now has been found using both 1) personality scores themselves (in student, applicant, and computer simulation samples) and 2) an external measure of verifiable faking such as the job desirability measure included in this study provides strong evidence that faking disproportionately increases an applicant’s chances of being selected when an organization uses non-cognitive measures for hiring purposes thereby affecting hiring decisions.

Beyond the important questions of whether applicants actually fake and whether faking affects hiring decisions, the current study made additional contributions to the faking literature by using an applied sample and a new external measure of faking to
investigate whether applicant faking patterns are job-specific. Do applicants fake in accordance with the perceived requirements of a specific job? Results from Hypothesis 3 and 4 suggest that this is indeed the case. Between- and within-subject comparisons indicated that applicants in each job group had higher scores on the job desirability measure that was specifically designed for their job. In the within-subjects analysis for Hypothesis 3, sales applicants had significantly higher scores on the sales scale than on the manager scale (using non-parametric tests) whereas manager applicants scored significantly higher on the manager scale than on the sales scale. This shows that when presented with bogus experience items for two different jobs, applicants chose to fake more on the bogus items that were specific to their job. Moreover, the between-subjects analyses for Hypothesis 4 revealed that sales applicants significantly outscored manager applicants on the sales scale and conversely, manager applicants outscored sales applicants on the manager scale, although this difference only approached statistical significance. The interpretation from these findings were further supported by results from the follow-up ANOVA, which confirmed the presence of a significant statistical interaction between job group membership and the version of the job desirability scale that participants responded to. Thus, not only did applicants fake most on bogus items relating to their desired job, but they also faked more on these items when compared to applicants who sought alternative jobs. Combined, the findings from Hypotheses 3 and 4 and the follow-up ANOVA provide strong evidence of job specificity in applicant faking patterns.

If applicants fake their responses to personality scales in a job-specific manner, and if they do the same on job experience questionnaires such as the job desirability scales
used in this study, then how much confidence can we hold in personality and non-cognitive test scores as a whole? The results of this study provide support both of these statements, and they have broad implications for the interpretation of these measures in a selection context. If applicants manipulate selection test items and succeed in creating profiles that are ideal for the specific job they are seeking, then the validity and utility of such items could be greatly reduced. In such a case, responses to non-cognitive measures could be considered to be entirely situation-specific such that individuals’ responses would vary greatly depending on the testing situation. Such situation specificity in response patterns could be attributed to faking. However, an alternative explanation could be that applicants describe themselves in a more favorable light when they consider themselves in a work-related context. In other words, they may describe their behavior as more conscientious, for example, when they are at work versus when they are at home. Markus and Wurf’s (1987) describe a body of research showing that individuals have a dynamic “working self-concept” that is multi-faceted, can change over time varies by context. Lord and Brown (2004), as cited recently by Johnson and Chang (2008), go further by stating that “in work contexts, phenomena such as leader behaviors, interpersonal interactions, and the work itself are capable of eliciting a particular working self-concept.” Moreover, individuals can display an average, or general, self-concept over time (Johnson & Chang, 2008). Hence, it is plausible that individuals can have an honest view of their personality that differs significantly depending on whether the context is in or outside of work. As applicants likely consider themselves from an “at work” context when they respond to non-cognitive measures, such as those used in this study, it is possible that the applicants in this study simply produced profiles which they felt honestly
described their personality at work (rather than outside of work). Whether or not this could be extended to their responses on the job desirability questions, which are much more biographical and verifiable in nature, is less likely, however.

Further research should investigate more closely how various factors related to the applicant setting (job requirements, how attractive the job is) and the applicant him/herself (personality traits, motivation to obtain a specific job) can elicit, modify, or even create a work-specific self-concept when applicants apply for a specific job. This dynamic self-concept, in turn, likely has important effects on job-specific faking behavior and more broadly, on individual differences in faking. Such a line of research would provide valuable insight into the faking processes used across different job application settings, and help identify the most effective ways to address each situation.

Hypothesis 5 was aimed at further investigating the nature of faking as a construct. It sought to clarify previous findings regarding the relationship between observed faking behavior and the impression management and self-deception sub-factors of social desirability (Carroll et al., 2004, Griffith et al., 2004). Unexpected findings emerged from these analyses. Contrary to expectations, self-deception correlated more with job desirability scores (i.e., faking behavior) than impression management. This difference was more pronounced for the sales group (.21 versus .12) than for the manager group (.35 versus .30). Carroll et al. (2004) found moderate correlations between impression management and their measure of job desirability but found negligible correlations between self-deception and job desirability. In the current study, however, job desirability correlated more strongly with self-deception for both sales and manager groups. This
leads one to the conclusion that the faking exhibited may be more subtle, and be more akin to putting one’s best foot forward rather than faking as much as possible.

Important implications arise from these findings. It was argued earlier that faking constitutes deliberate deception or misrepresentation of oneself, which is more similar to the impression management aspect of social desirability, whereas the self-deception aspect involves a more subconscious and unintentional form of misrepresentation. Based on this premise, a measure of deliberate faking such as the job desirability scales used in this study should correlate with impression management but should not correlate with self-deception. Contrary to this, the findings from Hypothesis 5 suggest that perhaps faking on the job desirability measures was somewhat subconscious in nature. As such, it should be noted that some researchers argue that impression management actually is the sign of a psychologically healthy person and that it is nearly impossible to distinguish faking from normal, “socialized” behavior because in everyday life, people are reinforced by society to make a favorable impression of themselves (Hogan et al., 2007; Hogan & Nicholson, 1998).

An examination of the inter-correlations between the study variables reveals interesting findings. Social desirability correlated very strongly with conscientiousness and competence (r’s ranging from .37 to .46 in the full sample), suggesting a strong relationship to true personality for both the impression management and self-deception sub-factors. In contrast, job desirability had a much weaker relationship with these two personality scales, with correlations ranging from only .11 to .15. Thus the job desirability measures measured variance that was distinct from conscientiousness (and competence, a highly similar construct). With respect to extroversion and gregariousness,
findings were different, however. Social desirability correlated only in the .17 to .38 range with these personality scales in the full sample. These correlations were very similar to the correlations that job desirability displayed with extroversion and competence (.17 to .30), indicating that job desirability did tap the extroversion construct to some extent – primarily in the sales group. Thus, job desirability successfully measures variance that is unique from personality, but this depends on the particular personality factor being considered, and possibly on the type of job applicants are applying for.

Although the positive correlation observed between self-deception and job desirability was unexpectedly significant, this finding should be interpreted within a broad perspective. The fact remains that all the correlations between job desirability and both the social desirability factors were statistically significant yet relatively small in magnitude. Although these correlations were not completely negligible, as in the case of Carroll et al. (2004), none of the correlations between job desirability, impression management, and self-deception exceeded a value of .35 in the sales and manager applicant samples, indicating that job desirability measured variance that is distinct from either social desirability factor. Stated another way, the maximum variance in job desirability scores that was accounted for by social desirability was only 12% based on the co-efficient of determination. Hence, job desirability measured variance that was unique from both personality (extroversion or conscientiousness) and social desirability (impression management or self-deception factors). It is thus a useful faking measure because 1) it is not closely tied to true personality score variance, which would unfairly flag individuals who are truly high on a desirable construct, and 2) it is different from
social desirability, which as discussed earlier, often has been ineffective in efforts to identify faking and minimize its potentially detrimental effects on validity.

Having established that job desirability taps unique variance that is unique from self-deception and impression management, an important question then becomes: which of these constructs is most useful in the detection of applicant faking? In examining the difference scores between applicants and incumbents on the social desirability and job desirability measures (see Tables 3c and 3d) it is evident that the job desirability measure was more sensitive to applicant-incumbent differences than the impression management measure. In the sales group, the job desirability $d$ value was .18 standard deviations compared to only .04 standard deviations for impression management. Although the difference of .18 for job desirability was relatively small and was not statistically significant, it was over four times more sensitive than impression management. In the manager group, the same pattern of results emerged, although both scales were more sensitive. The manager job desirability scale $d$ value was .57 standard deviations compared to only .28 standard deviations for impression management. This finding indicates that job desirability was highly sensitive to manager applicant-incumbent differences, as it represents an effect size of over one half standard deviations, and again shows that it is more sensitive to faking than impression management. In the case of managers, it was twice as sensitive as impression management.

Given the fact that impression management scales have been by far the most widely used faking measures in both laboratory and field studies alike, this is a finding of great consequence. If the consulting companies who provided data for this study had used the impression management scales to identify potential faking, little or no value would
have resulted from using the impression management scale in the sales group, as applicants scored only .04 standard deviations higher than incumbents. This represents essentially no difference between applicants and incumbents. As for the manager group, some value would have resulted (slightly over one quarter standard deviations difference) but it would have been trivial in comparison with the potential value of the job desirability scale, which exhibited twice the magnitude in effect size across applicants and incumbents. The impression management scale included in this study is one of the most widely researched instruments used to measure faking, which is alarming, given its low sensitivity to applicant-incumbent differences in the current study. By comparison, the job desirability measure developed for this study proved to be much more valuable in detecting applicant faking for both sales and manager jobs.

What is even more surprising about the small effect sizes found for the impression management scale is the fact that the self-deception scale actually displayed consistently higher effect sizes. For the sales group, the self-deception d value was .22 (.04 for impression management) and for the manager group, the self-deception d value was .41 (.28 for impression management). Based on Paulhus’ (1991) definition of self-deception as a subconscious misrepresentation of oneself that is linked to one’s true personality, one would conclude from these results that applicant-incumbent differences do not reflect a conscious effort to fake; rather they are indicative of unconscious and unintentional tendency to represent oneself in a favorable light. However, the sales job desirability effect size was almost as large as the self-deception effect size while the manager job desirability effect size was double the size of the self-deception effect size. The items in these job desirability measures constitute a much more verifiable measure of faking –
responding “yes” to these items should involve a higher level of deliberate falsification in comparison to elevating one’s score on the self-deception scale items. Thus, the observed applicant-incumbent differences could be attributed to deliberate deception based on the job desirability effect sizes, yet also attributed to unconscious misrepresentation based on the self-deception effect sizes, particularly for the sales group.

Further research is needed to understand these findings more clearly. One interpretation could be that faking is largely deliberate, and that applicants elevate their scores on the self-deception items consciously, rather than subconsciously. Alternatively, any elevation in applicant scores could be interpreted as an unconscious aspect of true personality, and as such, endorsement of job desirability items would be considered as unintentional. An example of this would be an applicant who indicated having a non-existent skill not because of a conscious effort to fake, but rather because the skill sounded like something they had learned in the past and appeared to be job-relevant. Perhaps subconsciously, the desire to be selected for a job would cause such an applicant to claim having experience with this skill even though they were not fully certain about this claim.

However, a third interpretation seems more plausible. Observed applicant-incumbent differences likely constitute both conscious and subconscious processes that result in some combination of deliberate deception and unintentional misrepresentation that ultimately raises applicant scores above typical incumbent score ranges. The correlations between personality, impression management, self-deception, and job desirability indicated applicant personality scores shared at least some variance with all three measures, particularly with self-deception and job desirability. As such, it is likely
that applicants engage in both deliberate and unintentional faking, and in a manner that can be either specific or non-specific to a given job. As noted previously, research has shown that there are individual differences in faking (Griffith et al., 2004; McFarland & Ryan, 2000). Thus applicants could vary widely in terms of 1) whether they engage in unconscious or deliberate faking behavior, and 2) whether they engage is some combination of the two.

The effect sizes found for the personality measures in this study were somewhat smaller than those found in previous research. In their meta-analysis of faking effect size estimates, Viswesvaran & Ones (1999) reported that the average faking effect size in simulated faking studies ranged from .48 to .65 standard deviations for Big Five personality measures. Within-subject faking simulations tended to result in higher effect sizes, ranging from .47 to .93. Zickar & Robie (1999) provided more accurate estimates using IRT methodology, which produced faking effect sizes that ranged from .63 to .82 under general instructions to fake but from .93 to 1.51 when participants were coached on how to fake. In line with these findings, an earlier study conducted by the author found that effect sizes for within-subject differences on a measure of the Big Five personality factors ranged from .34 to as high as 1.17 when comparing responses under instructions to fake versus instructions to respond honestly.

As shown in Tables 3a and 3b, the $d$ values for the four personality scales used in this study were somewhat lower, ranging only from .27 to .54. These findings are in line with those reported by Birkeland et al. (2006), who found that applicant-incumbent effect sizes in applied settings ranged from .11 to .45 depending on the Big Five scale in question. Hence findings from the current study further confirm that in actual selection
settings, applicants do tend to score significantly higher than applicants, but they do not fake as much as individuals who are instructed to fake in a laboratory setting. Some researchers (Birkeland et al., 2006; Viswesvaran & Ones, 1999) have attributed this discrepancy between applied and laboratory faking behavior to the fact that instructions to fake induce a level of faking that is artificially high whereas faking in selection settings is not induced or manipulated by an experimenter; rather it is a decision taken on the part of applicants, who likely differ in their motivation and intent to fake. Applicants are also subject to the possibility that they could be punished for faking by being removed from consideration for the job they seek. In fact, many assessments in use today display some sort of warning or statement that either discourages dishonest responding or clearly states that individuals who respond dishonestly will not be considered for a position with the company. In light of this, it is little wonder that applicants do not inflate their scores to the extent that is usually observed in laboratory settings where participants are asked to fake as if they are an applicant, or to enhance their resulting profile as much as possible. Thus the findings of this study further confirm that applicants do not elevate their scores to the maximum extent possible but rather raise their scores in a more subtle manner, and as shown by the correlations between personality, job desirability and self-deception, they may do so in both conscious and unconscious ways.

If applicants do not fake as much as laboratory participants, however, is it possible to accurately and reliably identify applicants who do fake? And if so, what is the best method to use? Findings from this study support the use of a job desirability measure for this purpose. The job desirability scale was clearly the most sensitive to applicant-incumbent differences in the case of the manager position. As discussed earlier, the
manager job desirability scale exhibited an effect size that was twice as large as the effect size for impression management. In addition, it was larger than the effect size for the self-deception measure. This supports previous findings by Miller & Tristan (2002) that job desirability measures are more sensitive to faking than social desirability measures in a laboratory setting under instructions to fake versus to respond honestly. Taken together, these studies provide both laboratory- and field-based support for the use of job desirability as opposed to social desirability as a method for identifying fakers, at least for managerial selection.

As for sales selection, the results of this study are not as conclusive, based on the fact that job desirability did not exhibit as high an effect size as it did for managers. However, the effect size of .18 was nonetheless higher than the .04 effect size for impression management, indicating again that the job desirability measure was more sensitive to faking than impression management. However, the sales job desirability effect size was much smaller in comparison with the manager job desirability effect size, and was also somewhat smaller than the self-deception effect size of .22. Thus, in the case of sales, job desirability would be more effective than impression management in terms of identifying fakers but not necessarily more effective than a self-deception scale.

Yet the results of the Hypothesis 6 analyses first must be considered before this conclusion is made. Hypothesis 6 was aimed at assessing whether incumbents scored significantly higher than zero on the job desirability measures. Mean scores that were significantly above zero would indicate that the faking measure could be incorrectly identifying participants as fakers. Stated differently, this finding could indicate that the job desirability measure was resulting in false positives, or participants categorized as
fakers who in actuality were honest and mistakenly endorsed bogus job experience items due to confusion or error (not due to dishonesty). Hypothesis 6 thus predicted that incumbents would not score significantly higher than zero on the job desirability scales, whereas applicants would score higher. Contrary to expectations, incumbents did score significantly higher than zero, although the effect size was not as large as it was for applicants. As discussed earlier in this section, one must consider the hypothesis that incumbents actually faked as well, rather than simply assume that they confused the bogus experience qualifications with ones that sounded similar to those which they actually did possess. Incumbents scored significantly higher than zero in both positions, although much more so in the case of sales. Put in perspective, the sales incumbent mean score (1.98) was nearly double that of manager incumbents (1.06). This unexpected high mean score for sales incumbents contributed to a non-significant mean score difference when it was compared to the sales applicant mean score (2.37).

Based on such high job desirability scores for sales incumbents, one could argue that the instrument did not necessarily result in a high rate of false positives, but rather that the incumbents in this job group engaged in faking even though they had little motivation to do so. As a consequence, it is premature to conclude that the job desirability measure was ineffective in identifying sales applicants who faked, or that it was less effective than the self-deception measure. It is likely that special attention must be given to faking measures designed for this particular job applicant population due to the nature of typical sales job requirements. Circumstances may dictate that if a job desirability measure is used with this population, specialized norms should be developed for sales applicants which take into account higher baseline incumbent mean scores and
identify only unusually high scoring individuals as fakers. In Hypothesis 2 of this study, a 95th percentile on the job desirability scales was used as the cutoff point to classify applicants as fakers. Based on this, only sales applicants with a score of 7 or higher were labeled as fakers whilst only manager applicants with a score of 6 or higher were labeled as fakers. A strategy such as this, which identified only extreme forms of faking on the job desirability measures, may be the most appropriate method for identifying fakers using bogus items.

From a standpoint of practicality and legal defensibility, this approach would be beneficial because it would provide a highly objective and verifiable means by which to classify someone as a faker; only those whom endorsed an abnormally high number of items would be flagged as dishonest applicants. In turn, a hiring organization could then more closely examine the hiring data for these individuals, or simply remove them from the hiring process based on deliberate falsification during the hiring process. More research on similar job desirability measures is needed in applicant settings to determine the ideal cutoff point at which an applicant can be reliably and accurately considered to be faking on job experience questions versus simply confusing bogus items with actual qualifications. Nevertheless, the results from this study provide a promising basis from which to extend this line of research further, and they provide a highly useful baseline estimate of how much applicants would fake on a bogus item faking measure under real selection conditions.

The last hypothesis examined in this study was Hypothesis 7, which predicted that years of job experience in a similar job would be significantly and positively correlated with job desirability scores for both job groups. This held true only for sales applicants,
however; there was no significant relationship between the job experience and job desirability scores in the manager group. This result is interesting in that job type could possibly mediate the relationship between job experience and faking. Specifically, sales applicants who were more experienced tended to fake more than less experienced sales applicants. For managers, however, there was no observed relationship between experience and faking.

One possible explanation for this effect could be that salespeople over time become more and more desensitized to embellishing the truth or being dishonest about personal experiences because they become accustomed to doing so in their day-to-day interactions with their contacts from prospect companies. This is plausible given that in order to be successful, salespeople must often seek to forge a common bond with prospects immediately during sales interactions. In order to accomplish this, a salesperson often engages in friendly conversation regarding various subject matters. In order to facilitate an interaction and increase chances of success during a sales call or meeting, a salesperson might claim to be knowledgeable about anything that their contact would ask about or mention even if they in fact lack knowledge about it. Over time, this process would condition a salesperson to become accustomed to this behavior and perceive it as increasingly acceptable. Ultimately, those with more years of sales experience might fake to a greater extent on selection measures simply because this behavior has become conditioned over time during their sales interactions. Further research should investigate the degree to which previous job experience correlates with faking behavior, and the degree to which experience in a specific type of job could lead to more job-specific patterns of faking.
Limitations of the Current Study

Field versus Laboratory Setting. As with any piece of research, there were many limitations to the current study. Due to the fact that this was a field study, an initial limitation was the loss of experimental control that one would normally have in a laboratory setting. As a result, it was not possible to control certain critical variables such as participant selection, administration, test motivation, and faking behavior. Participants were actual job applicants and employees who underwent the assessments administered by each of the two aforementioned consulting firms. Hence it was not possible to randomly assign participants to different conditions (employment status and job type) of the study. This aspect of the design also precluded the level of control over the administration of all questionnaires. Although all participants were given the same instructions immediately before they began the questionnaires for this study, there was no control over the instructions that participants were given at the beginning of the consulting firms’ assessments. Each consulting firm’s assessment was administered first, and each began with its own assessment instructions. Upon completing the firm’s assessments, participants then were directed immediately to the section containing the measures for this study, which commenced with their own set of instructions. Therefore, there were some differences in the instructions and the assessments that participants completed depending on which firm they participated through. These differences would have had a limited effect if any, however, because the second firm provided assessment data for only a small subset of the sample, a portion of the sales incumbent sample. Consequently, this firm provided no applicant data and no manager incumbent data.
It should be noted that the instructions of both consulting firms and the instructions at the start of this study’s questionnaires included a brief warning to respond as honestly as possible (see Appendix D). None of these warnings actually warned of possible repercussions that participants might face if they were identified as fakers, however. For example, both the consulting firms’ warnings simply instructed participants that they should respond as honestly as possible and suggested that they respond to items based on their initial reaction or impulse rather than try to think about what the most appropriate response might be. The only significant difference between the two firm’s warning statements is that the primary firm which provided most of the data went further by indicating that their assessment included validity scales which could detect faking. No reference was ever made to what could happen if a participant was actually identified as having responded dishonestly. These two sets of instructions were fairly consistent with those used for the questionnaire in this study, which also asked participants to describe themselves as honestly as possible. Thus, in spite of the fact that participants had to view different instructions that were outside of the scope of this study because of the firms’ assessment processes, all of these instruction sets were fairly consistent in nature and all included a succinct recommendation to respond honestly.

An issue that could be of greater concern is whether the warnings included in all of the instructions affected participants likelihood and extent of faking on all the measures in this study. Research has shown warnings can reduce the extent to which individuals fake (Dwight & Donovan, 2003). As such, it is possible that the subtle warnings described here may have suppressed faking levels to some extent. However, this concern is somewhat mitigated by the fact that these warnings were provided to all participants
(despite slight differences in their wording). Moreover, many psychometric instruments used today for selection and development purposes usually include some form of faking warning to test takers, thus the warnings included here simply provided participants with a highly realistic assessment experience which they would likely encounter in any other hiring or employee development scenario. Ultimately this makes the results of this study more replicable in other field settings.

Another limitation that resulted from the field setting aspect of this study was the lack of control over the actual testing conditions experienced by each participant. Participants completed the assessment on their own time in the location of their choice, thus introducing unwanted variability into the test setting. This could have affected scores to the extent that participants were distracted or to the extent that they experienced any technical difficulties (e.g., internet connection loss, issues with computer performance) or took pauses during the assessment itself.

Testing motivation also was outside the control of this study. As such it was impossible to induce a controlled motivation to fake or respond honestly, as is the case in most laboratory-based faking studies. This could introduce potential unwanted variability in personality and faking scores because individuals likely varied in their motivation to create a desirable or job-specific impression. Given that the overall goal of faking research ultimately lies in finding the best way to address applicant faking in actual selection settings, however, this limitation also serves as one of the principal strengths of this study: the fact that it was carried out in a real selection setting where motivation to fake did not need to be manipulated. An essential premise of faking research is that an applicant setting inherently induces a motivation to fake in order to increase chances of
one being hired for a job. All extant personality faking research is based on this premise (although there debate about whether this actually results in deliberate deception or unwanted variance in personality scores). Therefore, it follows that applicants in this study should have been motivated to score as highly as possible on the measures they completed and that this motivation was an inherent part of the applicant condition in this study. Whether or not the incumbent participants in this study felt little motivation to fake is a different matter which was addressed earlier in this section. Mean score levels for incumbents on the job desirability scales indicated that they may have engaged in some level of faking, although to a lesser extent than the applicants.

Lastly, a related limitation due to the field setting in this study was the lack of control over faking behavior itself. It was not possible to manipulate the extent to which individuals actually increased their scores or the pattern of their responses. With regards to the extent of faking, the results replicated various earlier studies which report that applicants tend to score anywhere between .10 and .50 standard deviations higher than incumbents (Birkeland et al., 2006). Thus it can be concluded that applicants faked their responses to a level that was expected. As for the pattern of their responses, this was not controllable either. For example, laboratory studies such as those by Miller and Tristan (2002) and Raymark et al. (2004) manipulated the type of job that participants were instructed to fake their responses for. This enabled the authors to test whether faking could be job-specific in nature on personality and faking measures. Although the field setting did not allow this level of experimental control, results still suggested that faking in an applied setting is to some extent job-specific, as evidenced by differences on the sales and manager job desirability across job types, and similar differences in personality
effect sizes across job types. This provided a much needed test of job-specific faking evidence in a field setting. The resulting findings provide complimentary support to previous laboratory research findings, thus lending increasing evidence that applicants fake with a specific type of job in mind, and that this affects the way in which they respond to non-cognitive psychometric measures.

*Non-normal Distribution of Data.* One of the challenges in the analysis was the non-normal distribution of the data. As described in the Results section, data for all personality scores was skewed and was indicative of kurtosis, thereby making it difficult to interpret results from parametric statistical tests, which are based on the assumption of normally distributed data. Violations of normality could result in underestimates of true relationships between variables, indicating that many of the correlations and mean score comparisons reported could be underestimates. To counter this limitation, non-parametric versions of all tests were conducted. These resulted in very similar findings for all analyses, thereby confirming results from the parametric tests and mitigating concerns about unwanted effects of non-normally distributed data on the interpretations made in this study. Due to the effect that faking tends to have on personality scores (skewed distributions, possible ceiling effects, etc.), further research should investigate new ways to address the challenges of non-normally distributed data that is often exhibited in both applied and laboratory faking studies.

Lastly, it should be mentioned that the design of this study unfortunately did not allow for the collection of performance data for the participants. Hence it was not possible to address one of the most critical of all questions in the faking literature: whether or not faking actually affects the criterion-related validity of non-cognitive
measures. Ultimately, faking researchers must seek an answer to this important question. If faking does not truly attenuate validity or adversely affect hiring decisions, it is of limited value as an area of research within Industrial/Organizational Psychology. It is not possible to assess whether those individuals who were identified as fakers in this study ultimately would have been poor performers or would have engaged in any counterproductive behaviors in an organization that might have hired them. As discussed in the literature review, research is inconclusive regarding this critical question. However, it can be said with certainty from the results of this study that applicants who faked on the job desirability measure did highly increase their chances of being hired based on personality scores in a top-down selection system because they were disproportionately represented at the top range of the applicant distribution. As the results for Hypothesis 2 indicated, the probability of a faker being selected increased linearly as the selection ratio became more selective. And although it cannot be said with certainty whether those who endorse bogus items are more likely to be poor performers, previous findings in applied settings by Pannone (1984) and Anderson et al. (1984) indicate that they indeed would be poor performers relative to those who answered these items honestly. Laboratory-based findings such as those reported by Mueller-Hanson et al. (2003) and Douglas et al. (1996) also support this hypothesis, which leads one to believe that the applicants classified as fakers in this study had a greater chance of being unsuccessful hires relative to other applicants.

Beyond the question of whether faking decays the criterion-related validity of a personality assessment, other practical and ethical implications would arise if an organization used an objective and verifiable faking measure such as the job desirability
scale used in this study. For instance, would an organization truly want to hire an individual who claimed to have several qualifications that were non-existent? Is there a meaningful difference between responding “yes” to one of the bogus items included in the current study and providing false information on an employment application or a résumé? In essence, the bogus qualifications that comprise the job desirability measures in this study can be likened easily to work-related qualifications such as a high school diploma, a training certificate, or previous experience performing any essential job skill. Any individual who has completed a job application in the United States within the past 20 years likely has encountered some form of formal disclaimer on that application stating not only that they could be disqualified from the hiring process if they supplied false information, but that their employment could be terminated at any point if the company was to discover that their application included false information. If companies have deemed it so unanimously unacceptable to allow false information on an employment application, falsification on a bogus job experience questionnaire (used as part of a selection process) should not be treated any differently.

Moreover, in this study, applicants were only classified as fakers if they had responded “yes” to 7 or more items in the sales group and 6 or more items in the manager group. In contrast, the Pannone (1984) study classified applicants as fakers if they endorsed only one item. The possibility of falsely classifying someone as a faker would have been much higher in that study, whereas in the current study, a rejection decision could be made with a much greater level of confidence because only extreme levels of faking on a verifiable faking measure would constitute faking. In other words, one would be much more confident that an individual was deliberately falsifying if they claimed to
have experience with six bogus qualifications versus only 1 because there would be a much lower probability that they had made six “honest” mistakes on their job application as opposed to making just one honest mistake. Thus, a hiring organization could benefit from using a job desirability measure such as the one used in this study. It would not only offer a viable alternative to traditional social desirability measures due to its lower correlations with personality variance but it would also provide a more verifiable and objective means of identifying dishonest individuals based on the fact that they had supplied false information during the application process.

**Summary**

The results of this study highlight the need for both researchers and practitioners alike to move beyond the use of social desirability scales as a measure of faking. This study confirmed what various faking studies have reported in the past – applicants fake their responses to non-cognitive tests in hiring situations. As previous research also has shown, there are rewards to this test taking approach; applicants who faked systematically increased their chances of being hired over non-fakers in a top-down selection system based on personality scores. More importantly, however, the findings from this study expanded upon the extant faking literature by testing an alternative measure of faking – a job desirability scale – in an applied setting. By comparing real sales and manager applicants’ scores on both versions of the job desirability scale and on multiple personality scales, it was possible to show that applicants fake in a job-specific manner by claiming to have job experience and personality traits that appear to be job relevant. Thus applicants fake more on personality and biodata items that measure job-relevant
constructs, as opposed to faking to the same extent on all items regardless of what construct they measure.

The current study also has implications for designs in applied faking research. Incumbent scores on the job desirability scales indicated that employees fake on non-cognitive measures at least to some extent, and particularly in sales jobs. Decades of applied faking research has made the assumption that incumbent scores represent honest scores. However, the results from this study suggest that this assumption may be faulty, and that careful scrutiny should be given when interpreting differences between applicant and incumbent scores on non-cognitive test scores. Moreover, the findings from this study suggest that in the case of sales jobs, applicant and incumbent scores may closely parallel each other.

Finally, the current study advances the understanding of faking and its relationship to social desirability. Results showed that although job desirability and social desirability were significantly correlated with each other, these correlations were generally small to moderate in size and were likely significant simply due to the large sample size. Therefore, it was shown that faking should not be synonymous with social desirability. Moreover, impression management displayed a smaller correlation with job desirability than self deception. This contradicts the theory that impression management scores constitute deliberate faking more so than do self-deception scores. Results suggest rather that faking variance reflects both overt/conscious and subtle/subconscious aspects of deception, in addition to variance from other constructs which are likely unaccounted for in this study. Taken together, these findings make a useful contribution to the existing
body of research on applicant faking and have wide implications for theory and practice on non-cognitive test use in selection settings
Table 1a

Demographic Frequencies for Study Sample Only

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>360</td>
<td>37.6</td>
</tr>
<tr>
<td>Male</td>
<td>570</td>
<td>59.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>28</td>
<td>2.9</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>93</td>
<td>9.7</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>7</td>
<td>0.7</td>
</tr>
<tr>
<td>Caucasian</td>
<td>677</td>
<td>70.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>52</td>
<td>5.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>52</td>
<td>5.4</td>
</tr>
<tr>
<td>Native American</td>
<td>6</td>
<td>0.6</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>Unknown</td>
<td>120</td>
<td>12.5</td>
</tr>
<tr>
<td><strong>Job Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales</td>
<td>762</td>
<td>79.5</td>
</tr>
<tr>
<td>Manager</td>
<td>196</td>
<td>20.5</td>
</tr>
<tr>
<td><strong>Job Type by Employment Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales Applicant</td>
<td>632</td>
<td>66.0</td>
</tr>
<tr>
<td>Sales Incumbent</td>
<td>130</td>
<td>13.5</td>
</tr>
<tr>
<td>Manager Applicant</td>
<td>112</td>
<td>11.7</td>
</tr>
<tr>
<td>Manager Incumbent</td>
<td>84</td>
<td>8.8</td>
</tr>
<tr>
<td><strong>Years Experience in Similar Job</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>118</td>
<td>12.3</td>
</tr>
<tr>
<td>1 or less</td>
<td>109</td>
<td>11.4</td>
</tr>
<tr>
<td>2-5</td>
<td>277</td>
<td>28.9</td>
</tr>
<tr>
<td>6-9</td>
<td>175</td>
<td>18.3</td>
</tr>
<tr>
<td>10 or more</td>
<td>199</td>
<td>20.8</td>
</tr>
<tr>
<td>Unknown</td>
<td>80</td>
<td>8.4</td>
</tr>
</tbody>
</table>
### Table 1b

**Demographic Frequencies for All Participants**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>782</td>
<td>43.9</td>
</tr>
<tr>
<td>Male</td>
<td>920</td>
<td>51.7</td>
</tr>
<tr>
<td>Unknown</td>
<td>78</td>
<td>4.4</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>224</td>
<td>12.6</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>143</td>
<td>8.0</td>
</tr>
<tr>
<td>Caucasian</td>
<td>1071</td>
<td>60.2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>112</td>
<td>6.3</td>
</tr>
<tr>
<td>Native American</td>
<td>10</td>
<td>0.6</td>
</tr>
<tr>
<td>Other</td>
<td>15</td>
<td>0.8</td>
</tr>
<tr>
<td>Unknown</td>
<td>205</td>
<td>11.5</td>
</tr>
<tr>
<td><strong>Job Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales</td>
<td>762</td>
<td>42.8</td>
</tr>
<tr>
<td>Manager</td>
<td>196</td>
<td>11.0</td>
</tr>
<tr>
<td>Professional/Technical</td>
<td>148</td>
<td>8.3</td>
</tr>
<tr>
<td>Sales Manager</td>
<td>144</td>
<td>8.1</td>
</tr>
<tr>
<td>Customer Service</td>
<td>62</td>
<td>3.5</td>
</tr>
<tr>
<td>Administrative</td>
<td>35</td>
<td>2.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>433</td>
<td>24.3</td>
</tr>
<tr>
<td><strong>Job Type by Employment Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales Applicant</td>
<td>632</td>
<td>35.5</td>
</tr>
<tr>
<td>Sales Incumbent</td>
<td>130</td>
<td>7.3</td>
</tr>
<tr>
<td>Manager Applicant</td>
<td>112</td>
<td>6.3</td>
</tr>
<tr>
<td>Manager Incumbent</td>
<td>84</td>
<td>4.7</td>
</tr>
<tr>
<td><strong>Years Experience in Similar Job</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>259</td>
<td>14.6</td>
</tr>
<tr>
<td>1 or less</td>
<td>217</td>
<td>12.2</td>
</tr>
<tr>
<td>2-5</td>
<td>534</td>
<td>30.0</td>
</tr>
<tr>
<td>6-9</td>
<td>327</td>
<td>18.4</td>
</tr>
<tr>
<td>10 or more</td>
<td>362</td>
<td>20.3</td>
</tr>
<tr>
<td>Unknown</td>
<td>81</td>
<td>4.6</td>
</tr>
</tbody>
</table>
### Table 2a
Means, Standard Deviations, and Correlation Matrix of Study Variables – Full Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>46.26</td>
<td>3.77</td>
<td>(.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>42.43</td>
<td>5.17</td>
<td>.45**</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>46.44</td>
<td>3.82</td>
<td>.82**</td>
<td>.51**</td>
<td>(.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>41.21</td>
<td>5.69</td>
<td>.39**</td>
<td>.94**</td>
<td>.45**</td>
<td>(.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>12.15</td>
<td>4.35</td>
<td>.41**</td>
<td>.20**</td>
<td>.37**</td>
<td>.17**</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>11.12</td>
<td>4.01</td>
<td>.46**</td>
<td>.38**</td>
<td>.45**</td>
<td>.36**</td>
<td>.60**</td>
<td>(.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>2.23</td>
<td>2.29</td>
<td>.15**</td>
<td>.27**</td>
<td>.14**</td>
<td>.30**</td>
<td>.12**</td>
<td>.24**</td>
<td>(.75)</td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>1.64</td>
<td>2.05</td>
<td>.13**</td>
<td>.17**</td>
<td>.11**</td>
<td>.18**</td>
<td>.12**</td>
<td>.20**</td>
<td>.63**</td>
<td>(.75)</td>
</tr>
</tbody>
</table>

Note: N = 1701 (N = 1699 for Job Desirability Manager and 1778 for Job Desirability Sales). *p < .05. **p < .01. SD = Standard Deviation.
Table 2b
Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Applicants and Incumbents

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>46.41</td>
<td>3.73</td>
<td>(.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>43.37</td>
<td>4.79</td>
<td>.46**</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>46.62</td>
<td>3.68</td>
<td>.81**</td>
<td>.55**</td>
<td>(.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>42.46</td>
<td>5.21</td>
<td>.40**</td>
<td>.94**</td>
<td>.49**</td>
<td>(.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>12.00</td>
<td>4.24</td>
<td>.39**</td>
<td>.19**</td>
<td>.33**</td>
<td>.18**</td>
<td>(.82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>11.40</td>
<td>3.78</td>
<td>.44**</td>
<td>.37**</td>
<td>.44**</td>
<td>.37**</td>
<td>.57**</td>
<td>(.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>2.30</td>
<td>2.26</td>
<td>.12**</td>
<td>.19**</td>
<td>.10**</td>
<td>.19**</td>
<td>.13**</td>
<td>.23**</td>
<td>(.74)</td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>1.51</td>
<td>1.95</td>
<td>.12**</td>
<td>.13**</td>
<td>.10**</td>
<td>.12**</td>
<td>.08**</td>
<td>.15**</td>
<td>.63**</td>
<td>(.74)</td>
</tr>
</tbody>
</table>

Note: N = 683. *p < .05. **p < .01. SD = Standard Deviation.
Table 2c  
*Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Applicants and Incumbents*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>45.93</td>
<td>3.58</td>
<td>.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>41.60</td>
<td>5.38</td>
<td>.38**</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>46.56</td>
<td>3.47</td>
<td>.79**</td>
<td>.42**</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>39.98</td>
<td>5.88</td>
<td>.32**</td>
<td>.95**</td>
<td>.38**</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>11.69</td>
<td>4.29</td>
<td>.40**</td>
<td>.18*</td>
<td>.34**</td>
<td>.17*</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>10.09</td>
<td>3.96</td>
<td>.43**</td>
<td>.30**</td>
<td>.44**</td>
<td>.28**</td>
<td>.57**</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>1.91</td>
<td>2.07</td>
<td>.21**</td>
<td>.31**</td>
<td>.17*</td>
<td>.33**</td>
<td>.23**</td>
<td>.29**</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>1.65</td>
<td>1.97</td>
<td>.17*</td>
<td>.15*</td>
<td>.18*</td>
<td>.15*</td>
<td>.26**</td>
<td>.31**</td>
<td>.61**</td>
<td>.73</td>
</tr>
</tbody>
</table>

Note: N = 196. *p < .05. **p < .01. SD = Standard Deviation.
Table 2d
*Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Applicants*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>46.50</td>
<td>3.57</td>
<td>.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>43.54</td>
<td>4.64</td>
<td>.45**</td>
<td>.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>46.71</td>
<td>3.58</td>
<td>.81**</td>
<td>.53**</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>42.61</td>
<td>5.10</td>
<td>.39**</td>
<td>.94**</td>
<td>.47**</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression</td>
<td>42.01</td>
<td>4.26</td>
<td>.39**</td>
<td>.18**</td>
<td>.33**</td>
<td>.17*</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>11.46</td>
<td>3.75</td>
<td>.45**</td>
<td>.37**</td>
<td>.45**</td>
<td>.38**</td>
<td>.60**</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability</td>
<td>2.37</td>
<td>2.28</td>
<td>.10*</td>
<td>.18**</td>
<td>.10*</td>
<td>.18**</td>
<td>.12**</td>
<td>.21**</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability</td>
<td>1.51</td>
<td>1.95</td>
<td>.11**</td>
<td>.14*</td>
<td>.09*</td>
<td>.13*</td>
<td>.10*</td>
<td>.16**</td>
<td>.63**</td>
<td>.74</td>
</tr>
</tbody>
</table>

Note: N = 632. *p < .05. **p < .01. SD = Standard Deviation.
Table 2e  
Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Applicants

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>46.75</td>
<td>3.09</td>
<td>.73</td>
<td>.40**</td>
<td>.34**</td>
<td>.39**</td>
<td>.38**</td>
<td>.37**</td>
<td>.36**</td>
<td>.35**</td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>41.89</td>
<td>5.26</td>
<td>.40**</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>47.14</td>
<td>2.73</td>
<td>.76**</td>
<td>.48**</td>
<td>.94**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>40.54</td>
<td>5.69</td>
<td>.76**</td>
<td>.68</td>
<td>.94**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
<td>.43**</td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>12.20</td>
<td>4.51</td>
<td>.39**</td>
<td>.22*</td>
<td>.35**</td>
<td>.24*</td>
<td>.24*</td>
<td>.24*</td>
<td>.24*</td>
<td>.24*</td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>10.78</td>
<td>3.90</td>
<td>.38**</td>
<td>.37**</td>
<td>.35**</td>
<td>.36**</td>
<td>.36**</td>
<td>.36**</td>
<td>.36**</td>
<td>.36**</td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>1.94</td>
<td>2.24</td>
<td>.23*</td>
<td>.33**</td>
<td>.28**</td>
<td>.35**</td>
<td>.35**</td>
<td>.35**</td>
<td>.35**</td>
<td>.35**</td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>2.10</td>
<td>2.19</td>
<td>.16*</td>
<td>.27*</td>
<td>.22*</td>
<td>.26**</td>
<td>.26**</td>
<td>.26**</td>
<td>.26**</td>
<td>.26**</td>
</tr>
</tbody>
</table>

Note: N = 112. *p < .05. **p < .01. SD = Standard Deviation.
Table 2f
Means, Standard Deviations, and Correlation Matrix of Study Variables – Sales Incumbents

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>45.29</td>
<td>5.29</td>
<td>(.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>41.31</td>
<td>6.10</td>
<td>.52** (.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>45.55</td>
<td>4.69</td>
<td>.81** .67** (.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>40.55</td>
<td>6.13</td>
<td>.46** .95** .61** (.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>11.84</td>
<td>4.03</td>
<td>.46** .28* .35* .31* (.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>10.61</td>
<td>4.09</td>
<td>.40** .29* .32* .27 .66** (.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>1.98</td>
<td>2.13</td>
<td>.19 .28* .12 .25 .21 .44** (.74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>1.45</td>
<td>1.92</td>
<td>.20 .04 .16 .03 -.15 .09 .57** (.74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: N = 51 except for JD Sales (N = 130). *p < .05. **p < .01. SD = Standard Deviation.
Table 2g

Means, Standard Deviations, and Correlation Matrix of Study Variables – Manager Incumbents

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness</td>
<td>44.85</td>
<td>3.91</td>
<td>(.74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extraversion</td>
<td>41.20</td>
<td>5.54</td>
<td>(.83)</td>
<td>.36**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>45.77</td>
<td>4.16</td>
<td>(.84)</td>
<td>.80**</td>
<td>.38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gregariousness</td>
<td>39.25</td>
<td>6.08</td>
<td>(.83)</td>
<td>.27*</td>
<td>.95**</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Impression Management</td>
<td>11.02</td>
<td>3.90</td>
<td>(.78)</td>
<td>.38**</td>
<td>.09</td>
<td>.31**</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Deception</td>
<td>9.18</td>
<td>3.86</td>
<td>(.76)</td>
<td>.42**</td>
<td>.20</td>
<td>.50**</td>
<td>.16</td>
<td>.56**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job Desirability Sales</td>
<td>1.87</td>
<td>1.85</td>
<td>(.65)</td>
<td>.20</td>
<td>.29**</td>
<td>.08</td>
<td>.30**</td>
<td>.06</td>
<td>.22*</td>
<td></td>
</tr>
<tr>
<td>8. Job Desirability Manager</td>
<td>1.06</td>
<td>1.44</td>
<td>(.63)</td>
<td>.05</td>
<td>-.11</td>
<td>.06</td>
<td>-.09</td>
<td>.09</td>
<td>.12</td>
<td>.51**</td>
</tr>
</tbody>
</table>

Note: N = 84. *p < .05. **p < .01. SD = Standard Deviation.
Table 3a

*Differences Between Applicant and Incumbent Personality Scores for Sales Job.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Applicant</th>
<th>Incumbent</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t (681)</th>
<th>Z</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>46.50</td>
<td>3.57</td>
<td>45.29</td>
<td>5.29</td>
<td>2.24*</td>
<td>-1.25</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>43.54</td>
<td>4.64</td>
<td>41.31</td>
<td>6.10</td>
<td>3.21**</td>
<td>-2.38*</td>
<td>.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>46.71</td>
<td>3.58</td>
<td>45.55</td>
<td>4.69</td>
<td>2.17*</td>
<td>-1.82</td>
<td>.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gregariousness</td>
<td>42.61</td>
<td>5.10</td>
<td>40.55</td>
<td>6.13</td>
<td>2.73**</td>
<td>-2.20*</td>
<td>.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.

Table 3b

*Differences Between Applicant and Incumbent Personality Scores for Manager Job.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Applicant</th>
<th>Incumbent</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t (194)</th>
<th>Z</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>46.75</td>
<td>3.09</td>
<td>44.85</td>
<td>3.91</td>
<td>3.81***</td>
<td>-3.78</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>41.89</td>
<td>5.26</td>
<td>41.20</td>
<td>5.54</td>
<td>0.89</td>
<td>-0.64</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>47.14</td>
<td>2.73</td>
<td>45.77</td>
<td>4.16</td>
<td>2.78**</td>
<td>-2.23</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gregariousness</td>
<td>40.54</td>
<td>5.69</td>
<td>39.25</td>
<td>6.08</td>
<td>1.52</td>
<td>-1.22</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.
Table 3c

*Differences Between Applicant and Incumbent Scores on Job Desirability and Social Desirability for Sales Job.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Applicant</th>
<th>Incumbent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Sales Job Desirability</td>
<td>2.37</td>
<td>2.28</td>
</tr>
<tr>
<td>Manager Job Desirability</td>
<td>1.51</td>
<td>1.95</td>
</tr>
<tr>
<td>Impression Management</td>
<td>12.01</td>
<td>4.26</td>
</tr>
<tr>
<td>Self-Deception</td>
<td>11.46</td>
<td>3.75</td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.

Table 3d

*Differences Between Applicant and Incumbent Scores on Job Desirability and Social Desirability for Manager Job.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Applicant</th>
<th>Incumbent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Sales Job Desirability</td>
<td>1.94</td>
<td>2.24</td>
</tr>
<tr>
<td>Manager Job Desirability</td>
<td>2.10</td>
<td>2.19</td>
</tr>
<tr>
<td>Impression Management</td>
<td>12.20</td>
<td>4.51</td>
</tr>
<tr>
<td>Self-Deception</td>
<td>10.78</td>
<td>3.90</td>
</tr>
</tbody>
</table>
Table 4a

Proportion of Sales Applicant Fakers Selected Under Varying Selection Ratios Based on Top-Down Ranking of Personality Scores.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Full Sample</th>
<th>.50</th>
<th>.25</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>.06</td>
<td>.07</td>
<td>.09</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.06</td>
<td>.07</td>
<td>.12</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>Competence</td>
<td>.06</td>
<td>.08</td>
<td>.09</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>Gregariousness</td>
<td>.06</td>
<td>.08</td>
<td>.09</td>
<td>.16</td>
<td>.20</td>
</tr>
</tbody>
</table>

N = 632. Note: Full Sample refers to full sample of sales applicants only.

Table 4b

Proportion of Manager Applicant Fakers Selected Under Varying Selection Ratios Based on Top-Down Ranking of Personality Scores.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Full Sample</th>
<th>.50</th>
<th>.25</th>
<th>.10</th>
<th>.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>.05</td>
<td>.07</td>
<td>.12</td>
<td>.17</td>
<td>.17</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.05</td>
<td>.05</td>
<td>.09</td>
<td>.18</td>
<td>.22</td>
</tr>
<tr>
<td>Competence</td>
<td>.05</td>
<td>.10</td>
<td>.12</td>
<td>.13</td>
<td>.13</td>
</tr>
<tr>
<td>Gregariousness</td>
<td>.05</td>
<td>.07</td>
<td>.09</td>
<td>.23</td>
<td>.29</td>
</tr>
</tbody>
</table>

N = 112. Note: Full Sample refers to full sample of manager applicants only.
Table 5

*Within-Subject Differences on Sales versus Manager Job Desirability Scores for Applicants.*

<table>
<thead>
<tr>
<th>Job Group</th>
<th>Sales JD</th>
<th>Manager JD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Sales</td>
<td>2.37</td>
<td>2.28</td>
</tr>
<tr>
<td>Manager</td>
<td>1.94</td>
<td>2.24</td>
</tr>
</tbody>
</table>

* *p < .05. **p < .01. ***p < .001. Note: N for Sales group = 632. N for Manager group = 111.*
Table 6

*Between-Subjects Differences on Sales versus Manager Job Desirability Scores for Applicants.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sales</th>
<th>Manager</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$t$ (741)</td>
<td>$Z$</td>
<td>$D$</td>
</tr>
<tr>
<td>Job Desirability – Sales</td>
<td>2.37</td>
<td>2.28</td>
<td>1.94</td>
<td>2.24</td>
<td>1.85</td>
<td>-2.33*</td>
<td>.19</td>
</tr>
<tr>
<td>Job Desirability – Manager</td>
<td>1.51</td>
<td>1.95</td>
<td>2.10</td>
<td>2.19</td>
<td>2.88**</td>
<td>-3.34**</td>
<td>-.29</td>
</tr>
</tbody>
</table>

Note: *$p < .05$. **$p < .01$. ***$p < .001$. 
Table 7

*Mixed-Design Analysis of Variance on Job Desirability Scores for Sales and Manager Scale Versions by Job Group.*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta_{p}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Group</td>
<td>1.15</td>
<td>1</td>
<td>1.15</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>Scale Version</td>
<td>23.04</td>
<td>1</td>
<td>23.04</td>
<td>13.82***</td>
<td>.02</td>
</tr>
<tr>
<td>Job Type X Scale Version</td>
<td>49.40</td>
<td>1</td>
<td>49.40</td>
<td>29.63***</td>
<td>.04</td>
</tr>
<tr>
<td>Subjects Scale Version</td>
<td>5544.36</td>
<td>741</td>
<td>7.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale Version X Subjects</td>
<td>1235.41</td>
<td>741</td>
<td>1.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.   **p < .01.   ***p < .001.
Table 8a

*Pearson’s r Correlations Between Job Desirability and Social Desirability Scale in Sales and Manager Applicant Samples.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>IM</th>
<th>SDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Desirability – Sales</td>
<td>632</td>
<td>.12*</td>
<td>.21**</td>
</tr>
<tr>
<td>Job Desirability – Manager</td>
<td>111</td>
<td>.30**</td>
<td>.35**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.

Table 8b

*Spearman’s Rho Correlations Between Job Desirability and Social Desirability Scale in Sales and Manager Applicant Samples.*

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>IM</th>
<th>SDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Desirability – Sales</td>
<td>632</td>
<td>.10*</td>
<td>.21***</td>
</tr>
<tr>
<td>Job Desirability – Manager</td>
<td>111</td>
<td>.21*</td>
<td>.33***</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.
### Table 9a

**One-Sample Tests on Sales Job Desirability Scores for Sales Incumbents and Applicants.**

<table>
<thead>
<tr>
<th>Job Group</th>
<th>M</th>
<th>SD</th>
<th>N</th>
<th>t</th>
<th>df</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent</td>
<td>1.98</td>
<td>2.13</td>
<td>130</td>
<td>10.60***</td>
<td>129</td>
<td>2.11***</td>
</tr>
<tr>
<td>Applicant</td>
<td>2.37</td>
<td>2.28</td>
<td>632</td>
<td>26.09***</td>
<td>631</td>
<td>4.48***</td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.

### Table 9b

**One-Sample Tests on Manager Job Desirability Scores for Manager Incumbents and Applicants.**

<table>
<thead>
<tr>
<th>Job Group</th>
<th>M</th>
<th>SD</th>
<th>N</th>
<th>t</th>
<th>df</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbent</td>
<td>1.06</td>
<td>1.44</td>
<td>84</td>
<td>6.73***</td>
<td>83</td>
<td>2.57***</td>
</tr>
<tr>
<td>Applicant</td>
<td>2.10</td>
<td>2.19</td>
<td>111</td>
<td>10.10***</td>
<td>110</td>
<td>2.07***</td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.
Table 10

Parametric and Non-Parametric Correlations Between Job Desirability and Job Experience in Sales and Manager Applicant Samples.

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Pearson’s r</th>
<th>Spearman’s rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Desirability – Sales</td>
<td>632</td>
<td>.26**</td>
<td>.26**</td>
</tr>
<tr>
<td>Job Desirability – Manager</td>
<td>111</td>
<td>.07</td>
<td>.02</td>
</tr>
</tbody>
</table>

*p < .05.  **p < .01.  ***p < .001.
### Table 11a

*Item Analysis of Sales Job Desirability Scale - Sales and Manager Applicant Samples.*

<table>
<thead>
<tr>
<th>Item</th>
<th>Sales Applicants</th>
<th>Manager Applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>I have experience with the alpha and beta call technique.</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>I have experience sublimating leads.</td>
<td>0.49</td>
<td>0.5</td>
</tr>
<tr>
<td>I have transposed a call log.</td>
<td>0.44</td>
<td>0.5</td>
</tr>
<tr>
<td>I have experience selling in blue markets.</td>
<td>0.16</td>
<td>0.36</td>
</tr>
<tr>
<td>I have taken Dr. Gerald Peterson’s <em>Personal Goal Structure</em> methods into account while developing my sales goals.</td>
<td>0.02</td>
<td>0.15</td>
</tr>
<tr>
<td>I have experience selling to ordinal distributors.</td>
<td>0.20</td>
<td>0.4</td>
</tr>
<tr>
<td>I have experience selling to bivariate customers.</td>
<td>0.21</td>
<td>0.41</td>
</tr>
<tr>
<td>I have experience doing telematrixing sales.</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>I have syntaxed quarterly or yearly sales profits.</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>I have syntonized revenue reports.</td>
<td>0.12</td>
<td>0.32</td>
</tr>
<tr>
<td>I have conducted a customer autogam.</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>I am aware of the American Marketing Organization’s <em>Salesperson’s Bill of Rights.</em></td>
<td>0.11</td>
<td>0.31</td>
</tr>
</tbody>
</table>
### Table 11b

**Item Analysis of Manager Job Desirability Scale - Sales and Manager Applicant Samples.**

<table>
<thead>
<tr>
<th>Item</th>
<th>Sales Applicants</th>
<th>Manager Applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have used the Critical Affects technique for performance appraisals.</td>
<td>0.09 0.29</td>
<td>0.14 0.34</td>
</tr>
<tr>
<td>I have experience in performance deviation.</td>
<td>0.26 0.44</td>
<td>0.38 0.49</td>
</tr>
<tr>
<td>I have applied a Managerial Sieve strategy to my management practices.</td>
<td>0.05 0.21</td>
<td>0.07 0.26</td>
</tr>
<tr>
<td>I have experience handling supply drifts.</td>
<td>0.14 0.35</td>
<td>0.21 0.41</td>
</tr>
<tr>
<td>I have had training in inventory covariance control.</td>
<td>0.17 0.38</td>
<td>0.15 0.36</td>
</tr>
<tr>
<td>I have experience with quality taxonomy groups.</td>
<td>0.02 0.15</td>
<td>0.03 0.16</td>
</tr>
<tr>
<td>I have used latency planning techniques as a manager.</td>
<td>0.05 0.22</td>
<td>0.11 0.31</td>
</tr>
<tr>
<td>I have calculated a liquifaction ratio.</td>
<td>0.12 0.32</td>
<td>0.14 0.34</td>
</tr>
<tr>
<td>I know how to calculate Return on Exponents (ROE).</td>
<td>0.11 0.31</td>
<td>0.18 0.39</td>
</tr>
<tr>
<td>I have proximated vendors as a manager.</td>
<td>0.15 0.36</td>
<td>0.28 0.45</td>
</tr>
<tr>
<td>I know how to incur cash in-flow.</td>
<td>0.34 0.47</td>
<td>0.41 0.49</td>
</tr>
<tr>
<td>I have read Daniel Goldberg’s book entitled <em>The Power of Emotional Leadership.</em></td>
<td>0.01 0.08</td>
<td>0.02 0.13</td>
</tr>
</tbody>
</table>
Figure 1

*Interaction Effects of Job Group X Job Desirability Scale Version on Job Desirability Scores.*
REFERENCES


Griffith, R. L., English, A., Yoshita, Y., Gujar, A., Monnot, M., Malm, T., & Graseck, M. (2004, April). Individual differences and applicant faking behavior: One of these applicants is not like the other. In N. D. Christiansen (Chair), Beyond Social Desirability in Research on Applicant Response Distortion. Symposium conducted at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.


incumbents. Paper presented at the 18th Annual Conference of the Society for
Industrial and Organizational Psychology, Orlando, FL.


Heron, A. (1956). The effects of real-life motivation on questionnaire response. Journal
of Applied Psychology, 40, 65-68.

impression management efforts on FIRO-B profiles. Journal of Consulting &
Clinical Psychology, 43, 269-275.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and

Tulsa, OK: Hogan Assessment Systems.

L. M. Hough (Eds.) Handbook of Industrial and Organizational Psychology,

American Psychologist, 43, 621-626.

Hogan, R. T., Hogan, J., & Roberts, B. W. (1996). Personality measurement and
employment decisions: Questions and answers. American Psychologist, 51, 469-
477.

Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in


James, L. R. (1999, April). *Use of Conditional Reasoning to distinguish between reliable and unreliable employees.* Paper presented at the 14th Annual Meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.


Appendix A

Goldberg’s (1999) IPIP Big-Five Factor Scales

Directions

In this section, there are phrases describing people's behaviors. Please use the rating scale below to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. Please read each statement carefully, and then select the response that best describes you.

Response Options

1 = Very Inaccurate  
2 = Moderately Inaccurate  
3 = Neither Inaccurate nor Accurate  
4 = Moderately Accurate  
5 = Very Accurate

*IPIP Conscientiousness Scale*

1. Am always prepared.  
2. Pay attention to details.  
3. Get chores done right away.  
4. Carry out my plans.  
5. Make plans and stick to them.  
7. Find it difficult to get down to work.  
8. Do just enough work to get by.  
9. Don't see things through.  
10. Shirk my duties.
IPIP Extraversion Scale

1. Feel comfortable around people.
2. Make friends easily.
3. Am skilled in handling social situations.
4. Am the life of the party.
5. Know how to captivate people.
6. Have little to say.
7. Keep in the background.
8. Would describe my experiences as somewhat dull.
9. Don't like to draw attention to myself.
10. Don't talk a lot.

IPIP Competence (analogue to Hogan's Managerial Potential) Scale

1. Come up with good solutions.
2. Complete tasks successfully.
3. Carry out my plans.
4. Accomplish a lot of work.
5. Get things done quickly.
6. Feel that my life lacks direction.
7. Am not sure where my life is going.
8. Hang around doing nothing.
9. Do just enough work to get by.
10. Mess things up.

IPIP Gregariousness (analogue to Hogan’s Sales Potential) Scale

1. Don't mind being the center of attention.
2. Feel comfortable around people.
3. Talk to a lot of different people at parties.
4. Am the life of the party.
5. Know how to captivate people.
6. Don't like to draw attention to myself.
7. Have little to say.
8. Dislike being the center of attention.
9. Would describe my experiences as somewhat dull.
10. Keep in the background.
Appendix B

Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991)

Directions

Use the rating scale below as a guide for the next set of questions. On a scale of 1 to 7, with 1 = “Not True” and 7 = “Very True,” indicate how much you agree with the following set of questions.

<--- 1 -------- 2 -------- 3 -------- 4 -------- 5 -------- 6 -------- 7 --->

NOT TRUE          SOMEWHAT TRUE

Impression Management Subscale

1. I sometimes tell lies if I have to.
2. I never cover up my mistakes.
3. There have been occasions when I have taken advantage of people.
4. I never swear.
5. I sometimes try to get even rather than forgive and forget.
6. I always obey laws, even if I’m unlikely to get caught.
7. I have said something bad about a friend behind his or her back.
8. When I hear people talking privately, I avoid listening.
9. I have received too much change from a salesperson without telling him or her.
10. I always declare everything at customs.
11. When I was young I sometimes stole things.
12. I have never dropped litter on the street.
13. I sometimes drive faster than the speed limit.
14. I never read sexy books or magazines.
15. I have done things that I don’t tell other people about.
16. I never take things that don’t belong to me.
17. I have taken sick-leave from work or school even though I wasn’t really sick.
18. I have never damaged a library book or store merchandise without reporting it.
19. I have some pretty awful habits.
20. I don’t gossip about other’s people’s business.

Self-Deception Subscale

1. My first impressions of people usually turn out to be right.
2. It would be hard for me to break any of my bad habits.
3. I don’t care to know what other people really think of me.
4. I have not always been honest with myself.
5. I always know why I like things.
6. When my emotions are aroused, it biases my thinking.
7. Once I’ve made up my mind, other people can seldom change my opinion.
8. I am not a safe driver when I exceed the speed limit.
9. I am fully in control of my own fate.
10. It’s hard for me to shut off a disturbing thought.
11. I never regret my decisions.
12. I sometimes lose out on things because I can’t make up my mind soon enough.
13. The reason I vote is because my vote can make a difference.
14. My parents were not always fair when they punished me.
15. I am a completely rational person.
16. I rarely appreciate criticism.
17. I am very confident of my judgments.
18. I have sometimes doubted my intelligence.
19. It’s all right with me if some people happen to dislike me.
20. I don’t always know the reasons why I do the things I do.
Appendix C

Job Desirability Scales

*Job Desirability Scale: Sales Version*

1. I have experience with the alpha and beta call technique.
2. I have experience sublimating leads.
3. I have transposed a call log.
4. I have experience selling in blue markets.
5. I have taken Dr. Gerald Peterson’s *Personal Goal Structure* methods into account while developing my sales goals.
6. I have experience selling to ordinal distributors.
7. I have experience selling to bivariate customers.
8. I have experience doing telematrixing sales.
9. I have syntaxed quarterly or yearly sales profits.
10. I have syntonized revenue reports.
11. I have conducted a customer autogam.
12. I am aware of the American Marketing Organization’s *Salesperson’s Bill of Rights*.

*Sales Version Distracter Items*

1. I have experience making cold calls.
2. I have experience qualifying prospects.
3. I have experience obtaining re-sales, or repeat sales.
4. I have experience obtaining customer referrals.
5. I have experience with system sales.
6. I have created projected sales reports.
7. I have calculated Return on Investment (ROI).
8. I have prepared expense reports.
Job Desirability Scale: Manager Version

1. I have used the Critical Affects technique for performance appraisals.
2. I have experience in performance deviation.
3. I have applied a Managerial Sieve strategy to my management practices.
4. I have experience handling supply drifts.
5. I have had training in inventory covariance control.
6. I have experience with quality taxonomy groups.
7. I have used latency planning techniques as a manager.
8. I have calculated a liquifaction ratio.
9. I know how to calculate Return on Exponents (ROE).
10. I have proximated vendors as a manager.
11. I know how to incur cash in-flow.

Manager Version Distracter Items

1. I have experience implementing Management By Objectives (MBO) principles.
2. I have experience implementing Total Quality Management (TQM) principles.
3. I have had training in Management Information Systems (MIS).
4. I have experience with supply chain management.
5. I have had training in ISO9000 Certification procedures.
6. I have assessed fixed and variable costs.
7. I know how to calculate Return on Total Assets (ROTA).
8. I have prepared close-out reports.
Appendix D

Assessment Instructions

*Instructions for Primary Consulting Firm*

Note: Since most business functions include some customer contact, the questions in the Sales Orientation Assessment are required for all positions. If you have never had direct sales experience, answer the questions according to how you believe they would apply if you were in sales. Note: a "No Opinion" response is acceptable and will not substantially affect your results.

**Instructions: Be sure to answer every required question.** If you skip any required questions, you will be prompted to complete them before you can go on to the next section.

- **If you change your mind about an answer,** scroll back up on the page and change your answer. Do this before you click on the "Click Here to Continue" button at the bottom of each page, because you won't be able to come back to that page of questions.

- **If you have to stop** before you have completed the assessment, don't worry: when you log back in to resume it will return you to where you need to continue. However, if you have finished a page of questions, those answers cannot be saved until you click on the button at the bottom of the page, and the following page of questions appears. If your browser times out before the next page appears (it will say "The page cannot be found", or something similar), refresh the screen using the Refresh and Retry commands if you are using Microsoft Internet Explorer as your browser. If you are using Netscape, use the Reload and OK commands.

- Give yourself every advantage: Your best bet is to be as honest as possible, and give your first-impulse answer. This survey is not timed, but please answer the statements as quickly as possible for the most accurate results. There are no "right" or "wrong" answers.

- **Caution:** Don't make the mistake of trying to out-think the questions and make yourself appear different from what is really true. There are validity scales built into this assessment.

- [Consulting firm’s name] do not report individual answers. Only a summarization and profile of your scores will be reported.
Instructions for Secondary Consulting Firm

Dear [FirstName] [LastName],

Congratulations, you have been chosen to complete the SalesPro assessment. This assessment asks you to answer questions about yourself how you approach your sales job. The information collected will be used by your organization to better understand their salesforce and to help make better hiring decisions.

Please complete this assessment at your earliest convenience. We appreciate you taking the time to participate in this assessment process and help your organization to improve and succeed.

The assessment will take approximately 60 to 90 minutes to complete. We suggest that you take the assessment in a quiet place where you can focus and concentrate. Please be honest in your responses. The more honest and accurate participants are when taking the assessment, the more accurate the results will be.

Before completing the assessment, please make sure that you have a high speed internet connection (e.g., Cable, DSL, T1). The assessment may not be able to run on a dial up connection.

To complete the assessment, click on the following link (link):

If there is no response when clicking the link above, please copy and paste the link into your browser and try again.

Then, follow these steps:

1. Log in by entering your email address (work email) and clicking “Continue”
2. Click on the link with your position title (HM Insurance: Consultant or Director)
3. Click on the “Launch Assessment” button and the assessment will soon begin

NOTE: if nothing happens after clicking “Launch Assessment”, you likely have a pop-up blocker still activated. A quick shortcut to bypass pop up blockers is to hold down the CONTROL key while you click on the “Launch Assessment” button, and continue holding the Control key for 2-3 seconds. This usually bypasses any blockers you may have turned on (e.g., Yahoo or Google task bars have their own pop up blockers).

To successfully view and complete the assessment, certain technical parameters must be met. If the assessment does not load after clicking the Launch Assessment button, please close all windows and re-enter the assessment. Please be sure to check your system for the requirements below.

TECHNICAL SPECIFICATIONS

1. Windows 98/NT/2000/XP
2. Mouse
3. Monitor screen resolution set to 800x600 with color quality set to high color (thousands of colors) or better
4. High-speed Internet Access (DSL connection acceptable; T1 or better is ideal)
5. Internet browser pop-up blockers need to be disabled (you can also hold Control key while launching assessment)
6. Internet Explorer 6 or higher
7. Flash Player Plug-in 7.0.19.0 or higher

If you continue to experience technical difficulties, please contact your organization’s HR representative.

Thank you.