

Wright State University
CORE Scholar

[Browse all Theses and Dissertations](#)

[Theses and Dissertations](#)

2012

Impression Management across Applicant and Incumbent Contexts: The Effect on Job Performance

Jenna Noelle Filipkowski
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

Filipkowski, Jenna Noelle, "Impression Management across Applicant and Incumbent Contexts: The Effect on Job Performance" (2012). *Browse all Theses and Dissertations*. 543.
https://corescholar.libraries.wright.edu/etd_all/543

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 library-corescholar@wright.edu.

Impression Management across Applicant and Incumbent Contexts:
The Effect on Job Performance

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

BY

JENNA NOELLE FILIPKOWSKI

B.S., Ursinus College, 2007

M.S., Wright State University, 2010

2012

Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

March 15, 2012

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Jenna Filipkowski ENTITLED, Impression Management across Applicant and Incumbent Contexts: The Effect on Job Performance, BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Corey E. Miller, Ph.D.
Dissertation Director

Scott N.J. Watamaniuk, Ph.D.
Director, HF-IO Psychology
Ph.D. Program

John M Flach, Ph.D.
Chair, Department of Psychology

Committee on Final Examination

Gary Burns, Ph.D.

David LaHuis, Ph.D.

Melissa L. Gruys, Ph.D., SPHR

Andrew Hsu, Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Filipkowski, Jenna N. Ph.D., Department of Psychology, Industrial and Organizational Psychology Program, Wright State University, 2012. Impression Management across Applicant and Incumbent Contexts: The Effect on Job Performance.

Social desirability (impression management) scales often accompany personality measures in selection to detect those who might be engaging in response distortion.

Applicants' personality scores may be corrected or eliminated based on scores from the impression management scale. My studies test the effectiveness and usefulness of having social desirability measures in personnel selection.

Study One examined whether social desirability (impression management) scales are able to detect faking behavior. The hypotheses were tested on an archival dataset of participants who took personality measures on two separate occasions as incumbents or applicants. Those identified by the faking indicator, who raised their scores beyond the standard error of measurement difference score, had higher extraversion scores in the applicant administration ($M = 19.16$; $SD = 1.83$) than as incumbent ($M = 13.88$; $SD = 5.27$). However, for the non-fakers extraversion scores from when the participants were incumbents are *higher* ($M = 19.00$; $SD = 2.13$) than their applicant scores ($M = 18.21$; $SD = 3.14$). Those identified as non-fakers while incumbents had such high scores that there was no opportunity to improve them and fake. The impression management scale was able to identify fakers on the personality measures who had opportunity to improve their scores.

For Study Two, I hypothesized that impression management is a necessary skill for salespeople, thus selecting people out or correcting their personality scores because of it results in worse people hired and lowers organizational performance. Impression management is not a necessary skill for salespeople; impression management does not have a relationship with weekly brokerage dollars ($r = -.07, p = .45$). Further analyses found that removing applicants or correcting their scores actually improved the job performance of the sample.

Keywords: Response Distortion, Faking, Personality, Impression Management, Personnel Selection.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
Personality Measures in Selection	2
Applicant Faking on Personality Measures	4
Individual Differences in Faking Behavior	8
The Prevalence of Applicant Faking	10
Methods for Detecting Fakers	18
Controlling Applicant Faking	23
Social Desirability Measures	29
Correcting Faking with Social Desirability Measures	32
Literature Review Conclusions	39
The Current Studies: Purpose and Contributions to the Literature	40
II. STUDY ONE	48
METHOD	48
Participants	48
Measures	49
Procedure	50
RESULTS	51
Tests of Hypotheses	51

III. STUDY TWO	57
METHOD	57
Participants	57
Measures	57
Procedure	58
RESULTS	59
Tests of Hypotheses	59
Research Questions	60
IV. DISCUSSION	65
Summary	65
Implications	68
Limitations	69
Conclusion	71
V. REFERENCES	94

LIST OF FIGURES

	Page
1. Extraversion Incumbent Scores.....	91
2. Extraversion Applicant Scores.....	92
3. Competitiveness Incumbent Scores.....	93
4. Competitiveness Applicant Scores.....	94

LIST OF TABLES

	Page
1. Study One: Score Differences between the Two Administrations.....	73
2. Study One: Descriptive Statistics and Intercorrelations for Group One.....	74
3. Study One: Descriptive Statistics and Intercorrelations for Group Two.....	75
4. Study One: Descriptive Statistics and Intercorrelations for Group Three.....	76
5. Study One: Descriptive Statistics and Intercorrelations for Groups Two and Three Combined.....	77
6. Study One: Non-fakers and Fakers for Groups Two and Three.....	78
7. Study One: Classification of Extraversion Faker and Non-Faker based on the Discriminant Function.....	79
8. Study One: Classification of Competitiveness Faker and Non-Faker based on the Discriminant Function.....	80
9. Study Two: Descriptive Statistics and Intercorrelations.....	81
10. Study Two: Weekly Sales Revenue for Sample, Selected by Extraversion Scores.....	82
11. Study Two: Average Supervisor Ratings for Sample, Selected by Extraversion Scores.....	83
12. Study Two: Weekly Sales Revenue for Sample, Selected by Competitiveness Scores.....	84
13. Study Two: Average Supervisor Ratings for Sample, Selected by Competitiveness Score.....	85
14. Study Two: Weekly Sales Revenue for Sample with Corrected Extraversion Scores	86

15. Study Two: Average Supervisor Ratings for Sample with Corrected Extraversion Scores.....	87
16. Study Two: Weekly Sales Revenue for Sample with Corrected Competitiveness Scores.....	88
17. Study Two: Average Supervisor Ratings for Sample with Corrected Competitiveness Scores.....	89

ACKNOWLEDGEMENTS

Foremost, I want to thank the loves of my life for their constant support and guidance throughout graduate school and my life. My parents, Mary Jean and Michael, and my brother Mikey have always been there to encourage me. I could feel how proud they are from 500 miles away. My fiancée, Jason's love lifted me up through the whole process. His skills in public speaking and writing provided me with valuable feedback traced all the way to my practice thesis defense presentation in 2009. I am grateful for our dogs, Lolly and Guinness who were lovable companions at home during long hours of work. I am also glad to have Erin, Cristina, and Zach in my cohort at Wright State. It was helpful to have people in my life who understand what graduate school entailed, especially Sara my high school friend in a different program in a faraway state whose commiserating gChat messages about school and life made me feel less alone. I want to thank all my friends and family who shaped my world beyond school.

Importantly, I want to thank my advisor Dr. Corey Miller whose knowledge and industry expertise were essential on this road to obtaining my doctorate. My entire committee, Dr. Corey Miller, Dr. Gary Burns, Dr. Melissa Gruys, and Dr. David LaHuis, were tremendous in their intelligent advice and reliability. I also want to thank my co-workers, Dr. James Killian and Dr. Christopher Holmes for providing feedback, encouragement, and opportunities.

I. INTRODUCTION

Selecting the correct person for a job opening is a major concern for organizations. In the applicant selection process, employers and applicants usually differ on their desired outcomes. Employers want to be certain that they are hiring the correct person for the job and that they spent their time and money wisely on testing and selection. Contrastingly, the majority of applicants want to be hired despite not being the best person in the applicant pool. Applicants present themselves as best as they can to achieve this outcome; therefore, some applicants may attempt to alter the way employers perceive them. One way applicants can do this is to distort their answers on personality measures that are used for selection, which is a concern to employers and Industrial/Organizational psychologists who design and administer these measures. If applicant faking on personality measures occurs and it negatively affects personnel selection, researchers need to determine what the best way is to deter or prevent faking behavior.

I/O psychologists have focused on the issue of applicant faking for the past 50 years. There are many ways to detect response distortion on personality measures, although none has been the silver bullet. They each have their advantages and disadvantages. The focus of this dissertation is on the oldest method of detecting applicant faking, social desirability measures. They have been used for decades alongside personality measures during selection in an attempt to identify respondents who are prone

to alter their scores. Social desirability measures are used in selection despite their effectiveness at identifying fakers. The focus of this paper is of the practical concerns of validity of these measures in a real-world sample.

This introduction will review the extensive literature on social desirability and complementary topics, including: the use and purpose of personality measures in the applicant selection process, theories of applicant faking on personality measures, the prevalence of applicant faking, the impact of applicant faking on the selection process, methods for detecting fakers and controlling applicant faking.

Personality Measures in Selection

One reason why employers often use personality measures in selection is that certain personality traits are valid predictors of job performance with minimal adverse impact against protected classes (Barrick, Mount, & Judge, 2001). A meta-analysis found that certain Big Five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability) predicted overall success in all jobs or specific performance criteria. Conscientiousness correlated with overall work performance ($\rho = .27$). Emotional stability was a valid predictor of work performance across jobs ($\rho = .13$) and was a valid predictor of teamwork ($\rho = .22$). Extraversion significantly correlated with teamwork ($\rho = .16$), training performance ($\rho = .28$), managerial performance ($\rho = .21$), and police officer performance ($\rho = .12$). Agreeableness and openness to experience had the lowest correlations across criteria and occupational groups; however, openness to experience predicted training proficiency ($\rho = .33$), and agreeableness predicted teamwork ($\rho = .34$) (Barrick, Mount, & Judge, 2001).

Personality traits are predictive of job performance and may be even more predictive than cognitive ability for particular jobs (Avis, Kudisch, & Fortunato, 2002). In a sample of customer-service employees, conscientiousness predicted job performance better than cognitive ability (Avis, Kudisch, & Fortunato, 2002). Being more conscientiousness was more important to successful job performance than higher cognitive ability was. Another meta-analysis of the Big Five personality inventory found that matching specific personality traits to specific criteria increases the predictive power of personality measures (Hogan & Holland, 2003). The estimated true validities from the Big Five ranged from $r = .34$ for agreeableness to $r = .43$ for emotional stability. Researchers also suggested a personality composite is a better predictor than individual personality traits in isolation (Barrick & Mount, 2005). In sum, personality traits are valid predictors for overall performance and specific performance criteria, and if the traits are considered jointly or matched to specific criteria, the validities improve substantially.

Besides being useful predictors of performance criteria, personality measures usually do not exhibit adverse impact in the selection process (Hogan, Hogan, & Roberts, 1996). Cognitive ability measures are the most successful predictors of job performance; however, they exhibit adverse impact, i.e., score differences for minority racial groups (Roth, Bevier, Bobko, Switzer, & Tyler, 2006). Personality measures exhibit less adverse impact than cognitive ability tests (Hough, Oswald, & Ployhart, 2001) and add incremental variance over cognitive ability measures (e.g., Avis, Kudisch, & Fortunato, 2002; Schmidt & Hunter, 1998).

Personality measures used in selection have many advantages. Certain traits are predictive of job performance, and the measures usually do not exhibit adverse impact. In

addition, researchers have advocated personality testing because longitudinal research has shown that personality predicts career success and other studies have shown personality to be related to less counterproductive work behaviors, turnover, tardiness, absenteeism, more citizenship behaviors, job satisfaction, task performance, and leadership effectiveness (Barrick & Mount, 2005). One disadvantage of personality testing is that respondents may not be truthful with their answers and may distort their responses to create a desirable image of themselves.

Applicant Faking on Personality Measures

There is no consensus among researchers on a definition of faking on personality measures (Zickar, Gibby, & Robie, 2004). The terms used for faking are response distortion, impression management, social desirability, self-enhancement, and claiming unlikely virtues (Griffith, Chmielowski, & Yoshita, 2007). Some researchers believe that faking is determined by a high score on a social desirability measure (e.g., Ones, Viswesvaran, & Reiss, 1996). However, faking behavior may be more complex than a score on a social desirability scale can account for, which will be discussed later in this paper. The definition of faking that I will use for the current research is an intentional form of response distortion used to create a favorable impression (Heggestad, Morrison, Reeve, & McCloy, 2006).

Researchers have not agreed on a universal definition of faking, nor have they agreed on a theoretical model of faking behavior. Five models will be highlighted here. Snell, Sydell, and Lueke (1999) developed an interactional model of faking, in which ability to fake and motivation to fake both influence successful faking. The interactional model includes the interaction of dispositional factors (cognitive ability and emotional

intelligence), experiential factors, and test characteristics that influence the ability to fake. Demographic factors, dispositional factors (impression management, integrity, Machiavellianism, manipulativeness, organizational delinquency, locus of control, and stage of moral development) and perceptual factors (others' behavior, others' attitudes, fairness, attitudes towards faking, expectations for success, and importance of outcome) influence motivation to fake (Snell et al., 1999).

Recently, Ziegler (2011) constructed a model of the conscious thought processes involved in faking. Ziegler conducted a qualitative study using a cognitive interview technique to ascertain the thought processes involved in faking personality measures. The results showed that fakers can be classified as extreme fakers or slight fakers, but neither type of faker faked all the items regardless of their content. Faking behavior described in his model had five cognitive stages: comprehension, importance classification, retrieval, judgment, and mapping. The results of his qualitative study supported this five stage model. A noteworthy finding that mirrors the other models is that there was an interaction between the person and situation in each of the stages. For example, test takers would evaluate an item based on its situational demand. If the test taker judged the item as unimportant with regard to the situation, no intentional faking would occur (Ziegler, 2011).

Tett, Anderson, Ho, Yang, Huang, and Hanvongse (2006) proposed a similar model of faking as the interaction of abilities, dispositions, and situations. They have not directly tested their model which is based on classic true score theory where an observed response on a personality item equals the targeted trait plus self-deception plus impression management plus the error term. Impression management is responding in

ways to make yourself look good, and self-deception occurs when you are unaware of your personality traits and you are unable to respond on the measure reflecting your true trait level (Paulhus, 1984). Tett et al. (2006) considered impression management and self-deceptive enhancement to be error terms that influence an observed personality score. In their interactive model the targeted trait, self-deception, and impression management are influenced by ability, targeted and non-targeted personality traits, and situational factors (Tett et. al, 2006).

Another applicant faking model uses James's Conditional Reasoning model to identify justification mechanisms for choosing responses on a personality test (Snell & Fluckinger, 2006). This applicant response model assumes that differences in justification mechanisms will moderate the validities of personality measures. Individual differences and situational factors are antecedents to justification mechanisms. Snell and Fluckinger (2006) did not directly test their model; however, they cited previous research that has supported these justification mechanisms (e.g., Bing, Whanger, Davison, & VanHook, 2004). For example, adding a frame of reference to personality items (such as "at work") increases the validity of the measure and produces mean changes in applicant responses (Snell & Fluckinger, 2006). Identifying justification mechanisms is an important step for developing new approaches for investigating faking, and by altering the justification mechanisms you can alter the respondent's answer choice on the personality measure.

McFarland and Ryan (2006) developed an integrated model of applicant faking behavior based on Ajzen's (1991) theory of planned behavior. McFarland and Ryan posited that one's attitude toward faking on personality tests, perceived social pressure to perform, and perceived behavioral control all influence an applicant's intention to fake.

The intention to fake is moderated by situational factors such as an incentive for doing well on the test and warnings not to fake. The influence of intention to fake on actual faking behavior is moderated by a responder's knowledge of the measured construct, self-monitoring, item transparency, and opportunity. McFarland and Ryan tested their theory of planned behavior to see if it predicted faking behaviors. Their study of 1,095 undergraduates found significant correlations between attitudes toward faking ($r = .64$), subjective norms toward faking ($r = .44$), perceived behavior control ($r = .47$) and intention to fake (McFarland & Ryan, 2006).

Ellingson (2011) drew conclusions about the faking models that they focus on factors that drive expectancy, instrumentality, and valence judgments. In the context of applicant faking, expectancy judgments concern the belief that one is capable of faking, instrumentality judgments concern the belief that faking is necessary, and valence judgments concern the belief that the opportunity is valued. She states three constructs that receive little research attention that may manifest valence in an applicant scenario. One, job desirability is the degree to which a job is wanted or needed. Two, marketability is the applicant's personal perception of his or her value to employers. Three, job search self-efficacy is a personal evaluation of one's ability to perform job search behaviors and obtain employment. She states that faking models need to look at the personal characteristics that distinguish fakers from non-fakers as well as distinguish between personal circumstances (i.e., job desirability) that may lead someone to fake (Ellingson, 2011). However, to detect faking in real-world applicant settings by using personal circumstance indicators would be impractical, time-consuming, and may produce adverse applicant reactions.

None of these faking models has become standard for faking research, but they all have many similarities. In particular, they describe faking behavior as done by applicant with certain characteristics, motivations, and cognitive abilities operating in a situation. Faking is likely to occur if the applicant has the ability and motivation to fake and faking a particular item is deemed important to being hired for the job. In other words, faking occurs when individual characteristics and situational demands produce systematic differences in a personality test score and these differences are not related to the attribute of interest.

Individual Differences in Faking Behavior

The models of applicant faking have a common tenant as they all see faking behavior as a complex interaction between personality, situations, and ability. This section goes into more detail about the individual differences between fakers and non-fakers. Researchers found two personality traits that are correlates of faking behavior (Griffith, Malm, English, Yoshita, & Gujar, 2006). Integrity and internal locus of control negatively correlated with faking. Those who are high in integrity and espoused the belief that they are in control of their life's outcome (internal locus of control) are less likely to fake. A surprising finding of this study is that two constructs commonly associated with faking did not correlate with the behavior. Impression management and self-deceptive enhancement did not positively correlate with faking behavior. Impression management is responding in ways to make yourself look good, and self-deception occurs when you are unaware of your personality traits and you are unable to respond on the measure reflecting your true trait level (Paulhus, 1984). They make up the two scales of Paulhus's (1984) Balanced Inventory of Socially Desirable Responding, a very common social

desirability measure. Griffith consistently makes the claim that social desirability measures are not effective proxies of applicant faking on personality measures (e.g., Griffith & Peterson, 2008). I will revisit this claim later in this manuscript.

Researchers examined the individual differences between fakers and non-fakers. Levashina, Morgeson, and Campion (2009) investigated the relationship between job applicant mental abilities and faking. In their study, job applicants ($N = 17,368$) for entry-level jobs completed a biodata measure and a cognitive ability test. A biodata instrument measures life experiences and typical behaviors in situations that are important to successful job performance. Embedded within the biodata measure were bogus items that assessed the candidate's familiarity with nonexistent events. Endorsing a bogus item was a proxy for faking. Applicants who endorsed the bogus items (faked) had higher scores on the biodata measure. The results showed that those with high levels of mental abilities were less likely to fake. However, for those who decided to fake, higher mental abilities helped applicants inflate their scores on the biodata measure over those with lower mental abilities (Levashina, Morgeson, & Campion, 2009).

Gammon and Griffith (2011) used a within-subjects design that had applicants take measures of conscientiousness, integrity, locus of control, self-efficacy, and counterproductive work behaviors. The same participants took the same measures six weeks later but for research purposes that were not related to their employment. Faking behavior was measured as a within-subject difference score between the two conditions. If the applicant score did not fall above the confidence interval (the standard error of measurement multiplied by 1.96), he or she was classified as a nonfaker. The results showed that fakers had lower integrity and self-efficacy (belief in one's abilities). Fakers

reported more counter-productive work behaviors and believed that the environment rather themselves controls life outcomes (high external locus of control). The results of this study show that fakers compared to nonfakers possess certain traits that may be harmful to the organization.

All of the above studies in this section on individual differences in faking behavior highlight certain traits that fakers possess. In sum, fakers are more likely to be low in integrity, internal locus of control, cognitive ability, and self-efficacy. These characteristics should not lead to successful job performance. I will explore the research on faking and job performance in a later section of this manuscript.

The Prevalence of Applicant Faking

Research has shown that between 30% and 50% of applicants elevate their scores on personality measures (Donovan, Dwight, & Hurtz, 2002) and 74% of applicants believe that other applicants fake (English, Griffith, Graseck, & Steelman, 2005). In 2011, Griffith and Converse summarized the research on the prevalence of applicant faking and found that 10% to 40% of applicants can be classified as fakers.

In addition, applicants may tailor their responses to what they believe the tester is seeking in a job applicant. In one study groups of students took personality tests with directions to answer as if they were presenting themselves as an ideal candidate for the jobs of a librarian, advertising executive, or banker (Furman, 1990). Different profiles emerged for each of the occupations; for example, the librarian was the most introverted (Furman, 1990). Another study found that students were able to fake a normative personality questionnaire that matched a profile of an ideal junior manager given by

actual mid-level managers or Human Resources managers from several organizations (Martin, Bowen, & Hunt, 2002).

These studies illustrate the fact that applicants may respond to personality measures in a way that they perceive the employer would want or that is stereotypic of the occupation. However, not all fakers are adept at doing this; research has shown that 20% of job applicants fake in the wrong direction (Burns & Christensen, 2006). In order to successfully fake a personality measure, meaning you score exactly how the administrator views a perfect applicant, the applicant must know exactly how to respond to achieve that perfect combination of scores. An applicant may attempt to respond as if he or she is an ideal candidate, but is unsuccessful at doing so. They may fake their responses in the wrong direction. Not all applicants know what the ideal applicant profile is, and no one can perfectly fake an entire personality profile (Hogan, 2005).

Research designs to study faking behavior. The differences in conceptualizations of faking result in the use of many types of research designs to determine the prevalence of faking behavior (see Mesmer-Magnus & Viswesvaran, 2006). In the laboratory, usually utilizing student samples, researchers determine if a personality measure is fakable by examining score differences from various instructional sets. Participants are often told to “fake-good”, meaning respond to the personality measure to make them look qualified for a job and to increase their chances of getting the job. Researchers instruct participants to answer honestly, meaning get a true reading of the trait level. Sometimes, researchers instruct participants to fake-bad, meaning respond in a way that makes a bad impression.

In laboratory studies, researchers examine the personality score differences from instructional set in either between-subjects designs or within-subjects designs. Within-subjects designs have greater statistical power than between-subject designs (Mesmer-Magnus & Viswesvaran, 2006). However, within-subjects designs are sensitive to threats to validity such as history, testing, and maturation. The order of administration for the within-subjects design for the honest and fake conditions does not affect the amount faking that occurs (Peterson, Griffith, Converse, & Gammon, 2011). Order effects for the honest/fake conditions are not a concern for laboratory research studies.

In contrast to laboratory studies, field studies examine faking behavior as score differences between applicants and incumbents in an organization. In addition, in field studies faking behavior can be determined by a high score on a social desirability measure. Another within-subjects design in an applied setting is to examine personality score differences of an individual as an applicant then as an employee of an organization.

Griffith, Chmielowski, and Yoshita (2007) examined the prevalence of applicant faking with a within-subjects design. Researchers gave applicants from two temporary employment agencies a customer service conscientiousness scale with their employment application materials. One month later, researchers mailed the same scale to the participants who now were employed ($N = 60$). They completed the scale under an honest condition and a fake-good condition. The mean scores for the three conditions were different [$F(2,59) = 43.32$, $p < 0.001$: applicant $M = 176.15$, $SD = 16.56$; honest $M = 164.92$, $SD = 18.35$; fake good $M = 191.79$, $SD = 27.23$]. Researchers also found that the rank ordering of applicants changed when response distortion occurred (Griffith et al., 2007).

The results of studies on faking. Meta-analyses of faking behavior in field and in laboratory studies found that faking does occur on personality measures (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Viswesvaran & Ones, 1999). In their meta-analysis of studies examining fake-good versus honest conditions, Viswesvaran and Ones (1999) found that individuals have the ability to fake on personality measures. On average, participants were able to raise their score almost one half standard deviation (Viswesvaran & Ones, 1999). Birkeland et al. (2006) conducted a meta-analysis of applicant versus incumbent studies and found that applicants' scores were higher on the Big Five personality traits than non-applicants' scores. The effect size was largest for conscientiousness and emotional stability ($d = .45$ and $d = .44$, respectively). They also found smaller mean differences between applicants and non-applicants, compared to the effect size reported in Viswesvaran and Ones' (1999) meta-analysis, possibly because of the use of non-laboratory samples. In laboratory studies, participants who are told to fake-good may exaggerate their response (Birkeland et al., 2006).

Not all researchers believe that faking occurs. Hough, Eaton, Dunnette, Kamp, and McCloy (1990) reported that faking might not occur in real-world applicant settings. They examined whether recently enlisted military recruits faked responses on a test after researchers told them that performance on the test affected decisions that were to be made about their careers. Results showed that these recruits scored lower than the other groups in the study. However, these recruits were not applicants (Hough et al., 1990). Further analysis of the data showed that 29% of the participants in the study were in fact faking (Rynes, 1993).

Similarly, researchers conducted studies on real applicants and found faking is not a problem (Hogan, Barrett, & Hogan, 2007). Hogan et al. (2007) used a within-subjects design of applicants to see if they changed their responses on the Hogan Personality Inventory, a measure of the Big Five personality traits. Applicants ($N= 5,266$) applied for a customer service job and the organization rejected them, and then six months later applicants reapplied and completed the same personality inventory. Results showed that less than 5% of the applicants improved their score. Researchers concluded that faking is not prevalent (Hogan et al., 2007). However, there is no way of knowing if these applicants were faking each time. In addition, in a recent study, Landers, Sackett, & Tuzinski (2010) suggested that those who are retesting after failure on a personality test and have low initial test scores are more likely to respond differently and improve their scores.

Ellingson, Sackett, and Connelly (2007) also used a within-subjects design and found faking is not the norm. Participants took personality measures as applicants then later as incumbents for developmental purposes. The researchers found that applicants engage in a limited amount of response distortion, as measured by effect size across contexts. The researchers noted some limitations in their study. Researchers used the California Personality Inventory (CPI) as the personality measure. The CPI is made of subtle items that may be less susceptible to intentional response distortion and it contains scales that measure general content not work-related content. Although 55% of the sample was managers, they did not report data for the CPI scales that would actually be used to select managers: Managerial Potential, Leadership Potential, and Work Orientation.

There were also several confounds in their study. There were changes in scores across development/development conditions. There was an effect for time; participants scored higher at time two than time one. There also was an effect for feedback. Participants scored higher the second time they took the CPI, often because they had received feedback. Ellingson et al. were forced to statistically correct for the effect of higher scores at time two, and the effect of feedback. It is important to note that Ellingson et al. used archival data, one of the limitations is that they were not involved with the data collection. Although there is a significant effect of time and feedback, Ellingson et al. were forced to statistically correct the data for confounding variables. This is not to say that Ellingson et al. made mistakes in analyzing the data, but we feel this is a limitation of their study that cannot be fully accounted for, even with statistical corrections. The researchers proposed that future research look at the impact of self-deception or unintentional distortion on the California Personality Inventory and use a broader sample of working individuals (Ellingson et al., 2007).

Researchers have come to seemingly contradictory conclusions. Some found evidence of faking in laboratory studies (e.g., Viswesvaran & Ones, 1999; Mueller-Hanson, Heggstad, & Thornton, 2003) and in field studies (e.g., Birkeland et al., 2006; Griffith, Chmielowski, & Yoshita, 2007), whereas others dismiss faking as a concern (e.g., Ellingson et al., 2007; Hogan et al., 2007). Laboratory studies usually measure faking as mean differences on personality scores between groups who have either been instructed to respond honestly or instructed to fake on the measure. Field studies, which are more generalizable to the applicant selection process, sometimes define faking as scoring high in social desirability.

Just as there is no universal operational definition for faking or an accepted theoretical model to guide research, there is no perfect way to detect the prevalence of faking. Some researchers believe that the fake-good format is a hypothetical exaggerated condition that is not representative of real-life settings (Smith & Ellingson, 2002). Although the research on the prevalence of faking is mixed, it may not be the most important concern to faking researchers. The impact of faking has on the selection process and hiring decisions may be more useful for understanding this process and for conducting future research.

The impact of faking on organizations. Employers spend money and time using personality measures in selection and they want an accurate score from the test taker. They do not want to hire someone based on their conscientiousness personality score and find out later that this person is not as conscientiousness as the test predicted. Researchers are concerned with the impact of applicant faking on the criterion-related or construct validity of the measure and the quality of personnel selection decisions made (Mueller-Hanson et al., 2003).

One study found that faking influences rank order in selection (Mueller-Hanson et al., 2003). The between-subjects design with students tested the effects of faking on selection. The researchers combined two groups of students (honest or incentive) into a single applicant pool and selected individuals for a hypothetical job based on varying selection ratios. The percentage of the honest group applicants were under-selected in all the selection ratios compared to the faking group. The researchers suggested the use of a select-out strategy to control for the effects of faking in real-world settings. The select-out strategy eliminates low scorers on personality measures from the applicant pool and

retains average to high scorers for future testing and interviews (Mueller-Hanson et al., 2003).

Another study found that using incumbent scores for cutoffs in selection costs the organization extra money in the selection process (Bott, O'Connell, Ramakrishnan, & Doverspike, 2007). They found no significant mean score differences between applicant and incumbents on a cognitive ability test ($d = .16$); however, mean scores on the personality measure were higher for the applicants than incumbents. Applicants were able to “fake” and raise their personality scores to higher levels. The researchers hypothetically set cut-off scores that then were applied to the personality test results of the incumbents and the applicants. The pass rates were much higher for the applicant group. If they used the pass rates from the applicant group, the organization would pay an extra \$61,300 in the selection process. Letting more people through increases the number of candidates to be interviewed (Bott et al., 2007). Researchers found that using multiple-predictors in selection such as a personality measure and a cognitive ability measure will reduce the amount of fakers hired (Peterson, Griffith, & Converse, 2009).

Researchers conducted a Monte Carlo investigation of the effects of faking on criterion-related validity of the personality measures (Komar, Brown, Komar, & Robie, 2008). To be effective predictors in selection personality measures need to keep their validity intact. They found that validity change is dependent on several parameters that vary across selection contexts. The parameters used in their study included magnitude of distortion, proportion who distort their responses, variability in the extent of faking, the faking-conscientiousness relationship, the faking-performance relationship, and the selection ratio (Komar et al., 2008).

Applicant faking on personality measures may have negative consequences for selection. However, some researchers believe that there may be some instances where successful faking on personality measures may be a good thing for certain jobs, e.g. sales jobs (Hogan, Hogan, & Roberts, 1996). Fakers may be low performers, high performers, or equal performers (see Peterson & Griffith, 2006). More research is needed to investigate the effects of faking on job performance.

Methods for Detecting Fakers

As seen in the section above, faking affects the selection process. Researchers have been addressing faking on non-cognitive measures since the 1930s (for a full review, see Zickar & Gibby, 2006). The most popular and oldest method for detecting fakers is the use of social desirability scales. The newer methods for faking detection are a verbal protocol analysis, reporting response latencies, bogus items, idiosyncratic item responses, Bayesian truth serum, item response theory, and structural equation modeling. The newer methods of faking detection will be discussed below followed by the research on social desirability measures.

The most direct method for detecting faking is using a verbal protocol analysis. A verbal protocol analysis requires participants to say whatever comes to mind when completing a task. Robie, Brown, and Beaty (2007) had 12 non-student participants with work experience complete a paper and pencil personality inventory as they verbalize their thoughts. Researchers told the participants that their personality scores would be compared to the job requirements of a job advertisement they read for a retail sales position and that the top three closest matches would each receive prizes. Participants' verbal responses showed that some did fake, and there were three classifications of

fakers: honest, slight fakers, and extreme fakers. Honest responders took less time completing the inventory and made fewer corrections of their responses (Robie et al., 2007).

Reporting response latencies of respondents is another method of detecting faking. However, the results of its effectiveness have been mixed. Some studies have found fakers have slower response times whereas other research has found fakers to respond faster (see Vasilopoulos, Reilly, and Leaman, 2000). Vasilopoulos et al. (2000) wanted to clarify the research on response latencies and proposed job familiarity as a moderator. Job familiarity is knowing what the job description is and what knowledge, skills, and abilities that the job requires. In their study, 116 students completed the Balanced Inventory of Desirable Responding Impression Management (BIDR-IM) (Paulhus, 1984) scale on a computer that captures the rating and the response latency. Researchers used the BIDR-IM to see if impression managers over report their desirable behaviors and under report undesirable behaviors. First, all participants completed a self-report honestly; then they were assigned into either an honest and low job familiarity condition, honest and high job familiarity condition, fake-good and low job familiarity condition, or a fake-good and high job familiarity condition. Participants completed the BIDR-IM and two scales that measured conscientiousness and emotional stability. Job familiarity moderated the relationship between response latency and impression management. Those told to fake good and had job familiarity (given a job description) had faster response times than those who did not have the job description (Vasilopoulos et al., 2000).

Using bogus items to detect fakers, Paulhus, Harma, Bruce, and Lysy (2003) conducted four studies on the over-claiming technique as a measure of self-enhancement. In the over-claiming technique, respondents rate their knowledge of various persons, things, or events. Twenty percent of the items on the measure are nonexistent. The researchers found that the over-claiming technique was valid even when respondents were warned about the foils and were asked to fake. The over-claiming technique showed convergent validity with other measures of self-enhancement and correlated with cognitive ability ($r = .52$) (Paulhus et al., 2003). Tristan (2009) used bogus items in a real applicant selection and found that applicants faked most on bogus items that were specific to the job they were applying, for either sales or manager.

Another technique for detecting faking is using idiosyncratic response patterns. Patterns could be detected when fakers were responding to items if the most desirable level of the trait was unclear, ambiguous, or debatable (Kuncel, Borneman, & Kiger, 2011). Kuncel and Borneman (2007) had 215 undergraduates complete the Goldberg Adjective markers, a multidimensional personality questionnaire, the Wonderlic, and the Balanced Inventory of Desirable Responding. The students completed the measures in two sessions: a honest condition or a faking condition (participants pretended to be applicants who really wanted the job). Researchers split the data so that each applicant's score contributed to either the honest condition or faking condition. The researchers computed the response distributions on the Goldberg Adjective Markers in both conditions. Researchers computed frequency distributions for the items with skewed distributions in both the honest and faking condition. They weighted the skewed items by how big the discrepancy of scores was between the honest and faking condition. A value

of 1 was assigned to the item if more people chose that item in the fake good condition and a value of -1 applied to items where more people in the honest condition chose that item. Moderate discrepancies were assigned .5 for higher responding in the fake good condition and -.5 in the honest condition. The scoring schemes used by the authors differentiated between scores in the honest versus faking conditions in cross-validation samples ($r = .45$ and $r = .67$) (Kuncel & Borneman, 2007).

Another complex but novel faking detection method is called Bayesian Truth Serum. This method requires two pieces of information from the test taker: the response to the item and the person's estimation of how many people in the population will respond in the same manner (Kuncel, Borneman, & Kiger, 2011). A formula gives high score to answers that surprisingly common, i.e., whose actual frequency exceeds their predicted frequency. Honest responses are given more weight. This approach is based on the assumption that people judge the frequency of the popularity of an item based on their own behavior, values, and preferences. If someone gives a response that is above the sample's collective predictive frequency, the item is considered more honest. This method was successful at detecting faking, but it has several drawbacks including that it requires more time and data analysis (Kuncel, Borneman, & Kiger, 2011).

Another newer method to detect faking involves the use of Item Response Theory (IRT). In one study using IRT, Zickar and Robie (1999) gave military recruits a personality inventory under one of three conditions: honest, fake-good, or fake-good with coaching. The researchers conducted IRT analyses to see if there were differences in option response functioning across the conditions. They used the changing persons model to measure theta shift between fakers and non-fakers. The thetas (ability level) between

fakers and honest respondents differed across the personality scales and conditions (Zickar & Robie, 1999).

Several studies have used IRT to identify differential responding between fakers and non-fakers (e.g., Holden & Book, 2009; Alarcon, 2011; O'Brien & Lohuis, 2011); however, the results of the studies depend somewhat on what types of models and cutoffs are used. Researchers have pointed out some problems with using IRT to detect faking. IRT assumes that previous items do not influence response to an item, but responses to survey items tend to be influenced by previous items (Mesmer-Magnus & Viswesvaran, 2006).

Finally, in another newer method, researchers treated faking as spurious measurement error caused by the interaction between context and person and used structural equation modeling to separate trait and faking variance (Ziegler & Buehner, 2009). They found controlling for faking reduced the intercorrelations among the Big Five personality traits and suggested that future research could use this model to detect faking. In addition, Ziegler and Buehner (2009) found that the personality variables had a relationship with the criterion while the faking did not. Structural equation modeling can be used to parse faking variance from trait variance in personality measures, but this is a complex practice for all applied settings.

In sum, the newer methods of detecting faking on personality test have been effective in identifying the behavior. However, most of the methods are used within a laboratory setting or may not be practical for real applicant settings. Instead of using controversial methods to detect faking, researchers have developed measures that are less susceptible to faking.

Controlling Applicant Faking

Various methods can detect fakers, and all have advantages and disadvantages. There are also various measures to control applicant faking, i.e., measures that are not easily faked. This section highlights the research on the methods and measures used to control for faking, including: using subtle items, item elaboration, time constraints, item randomization, warnings, forced-choice measures, Conditional Reasoning tests, and Implicit Association Tests.

The use of subtle items was one of the first methods used by researchers (Mesmer-Magnus & Viswesvaran, 2006). With subtle items, the administrator disguises his or her intent. Researchers have found subtle items to be less valid than obvious items (Hough et al., 2006). Item elaboration is another method for controlling faking. Participants who elaborated on their answers on the elaborated form of the biodata composite tended to have lower scores than the non-elaborated answers group (Schmitt, Oswald, Kim, Gillespie, Ramsay, & Yoo, 2003). Item elaboration has several drawbacks. It is labor and resource intensive, requiring follow-up interviews and more time to assess the measure. It is also taxing on individuals' mental abilities and handwriting.

Requiring respondents to fill out a personality measure under time constraints also taxes their abilities. Researchers found that time constraints reduced faking on a personality measure; however, this effect was only seen in individuals low in cognitive ability (Komar, Komar, Robie, & Tagger, 2010). In another study, researchers found that unproctored and speeded personality measures produced mean scores differences between applicants and incumbents that were similar to those reported for proctored tests (Arthur, Glaze, Villado, & Taylor, 2009). Thus adding a time constraint to a personality

measure may not be the best way to reduce faking, since there are only differences between speeded and not for those low in cognitive ability, which makes personality measures cognitively loaded.

Another method for controlling applicant faking is to randomize personality items throughout a test. Test administrators place similar constructs throughout the test rather than group them together. One study found that the grouped construct format was more fakable than randomized format for personality scales measuring neuroticism and conscientiousness (McFarland, Ryan, & Ellis, 2002).

One method given a lot of support is the use of warnings to control faking behavior. Dwight and Donovan (2003) examined the effectiveness of warning applicants not to fake on personality measure and three different types of warnings not to fake. Previous warning research found an average weighted mean effect of .23, i.e., warnings had a small effect on responses. Applicants warned not to fake have lower predictor scores than unwarned applicants. The type of warning, either identification-warnings (“the test contains items to identify fakers”) or consequences-warnings (“we will find out that you are faking, and you get in trouble”) influences the effectiveness of the warnings. Meta-analytic findings showed that identification-warnings had a negligible effect ($d = .01$), consequence-warnings had a moderate effect ($d = .30$), and both types of warnings together had a small effect ($d = .25$) (Dwight & Donovan, 2003).

McFarland and Ryan (2006) found warning applicants not to fake during the test had a direct effect on intention to fake and actual faking behavior. In another study, researchers found that these warnings deterred blatant extreme responding (faking) in a sample of applicants who were retesting after failure and may have been coached

(Landers, Sackett, & Tuzinski, 2010). Warnings may not be effective for all respondents; they may cue risk-takers to attempt faking (Tett et al., 2006). In addition, they may not be useful in the long run as test-takers catch on to the method (Zickar & Gibby, 2006). Warnings may be best used as a supplement to other types of faking detection measures (Tett et al., 2006).

Forced-choice measures are another method for controlling faking. Forced-choice items are unique in that all choices are socially desirable but not all are valid. The first forced-choice format, the Kuder Preference Record, was developed in the late 1930s (Zickar & Gibby, 2006). Christiansen, Burns, and Montgomery (2005) found that the forced-choice method was a better predictor of supervisors' ratings of performance than the single-stimulus method. They also found that cognitive ability positively correlated with successfully faking forced-choice items ($r = .25$ forced-choice and $r = .15$ single-stimulus) (Christiansen et al., 2005). It is not preferable in selection to have personality measures correlate highly with cognitive ability; adverse impact against protected groups could become a concern.

Some researchers have noted the other limitations of forced-choice measures (Converse, Oswald, Imus, Hedricks, Roy, & Butera, 2006). Some forced-choice items are ipsative, measuring intra-individual differences and not inter-individual. Inter-individual differences can be assessed with partially ipsative measures: ones that allow test-takers to partially rank order item alternatives, ones that have differing number of items, or ones with different scoring for responders with different characteristics. Heggstad et al. (2006) found that score comparisons of Likert type personality measures and forced-choice IPIP scales were similar, and modest effects occurred in the rank ordering of

individuals with either of the two types of measures. Waters (1965) found that respondents can successfully fake responses on forced-choice measures (cited in Hough et al., 1990). Some research has shown that applicants prefer traditional Likert items to the forced-choice format (Converse et al., 2006). Forced-choice formats are time consuming to create and may produce frustration for the test-takers (Zickar & Gibby, 2006).

Another method for controlling for faking on personality measures is the use of a Conditional Reasoning Test. Conditional Reasoning Tests assess the latent motives of an individual while disguising itself as a logical reasoning problem-solving test. LeBreton, Barksdale, Robin, and Lawrence (2007) examined Conditional Reasoning Tests to see if they were prone to response distortion. The researchers had the assumption that while keeping the indirect nature of the test intact, the Conditional Reasoning Test would not be susceptible to faking or social desirability bias. LeBreton et al. (2007) tested their assumption in three studies. Study one, comprised of undergraduate students, found those in the experimental group (those who were disclosed on the nature of the test) had higher mean scores than those in the control group, but those who were told to fake rather than find the most logical answer had the highest mean scores (control $M = 3.62$, $SD = 2.02$; disclose-logic $M = 4.49$ $SD = 2.51$; disclose-fake $M = 17.82$, $SD = 3.83$). Study two used a within-subjects design with the indirect measurement condition and researchers did not find significant mean differences between the control and applicant conditions. Study three examined the scores of job applicants, job incumbents, and undergraduates and the no significant mean differences were found ($M = 3.32$, $SD = 2.15$; $M = 3.30$, $SD = 2.13$; $M = 3.55$, $SD = 2.02$, respectively). Thus, Conditional Reasoning Tests can be faked

successfully if respondents are aware that the test was measuring their personality and was not problem solving test (LeBreton et al., 2007).

There have not been many Conditional Reasoning Tests created. James, McIntyre, Glisson, Bowler, and Mitchell (2004) created the first Conditional Reasoning Test. The Conditional Reasoning Aggression Test has reliability of $\alpha = .76$ and an average validity of $r = .44$. Other Conditional Reasoning measures of team orientation, anti-social personality, and social bias are being developed (LeBreton et al., 2007). Researchers have suggested the use of Conditional Reasoning Tests to combat faking behavior (Morgeson et al., 2007). However, the development of Conditional Reasoning Tests is onerous (Robie et al., 2007).

Besides being very difficult to develop Conditional Reasoning test have displayed another drawback, they show adverse impact as a selection measure towards minorities. Tristan, Miller, and Leasher (2003) found in a within-subjects design that the Conditional Reasoning test was less fakable in a fake-good condition; however, the fake-good condition showed adverse impact with African-Americans (effect size honest condition $d = .48$ and faking condition $d = .55$). The effect size for the Conditional Reasoning test under the faking condition was larger than two other personality measures of conscientiousness (NEO-FF $d = .44$ and Conscientiousness Biodata Questionnaire $d = .44$). The researchers proposed three explanations for the adverse impact: the reading level of the conditional reasoning measure, cultural bias, and stereotype threat (Tristan et al., 2003).

One of the newest measures to control for faking behavior is with the Implicit Association Test (IAT). Research on the Implicit Association Test has begun in the

discipline of Industrial/Organizational Psychology, admittedly with controversy (Landy, 2008). Implicit measures are distinct from explicit (self-report) measures; implicit measures do not require self-insight (Robinson & Neighbors, 2006). Explicit measures based on self-report are a measures of an individual's subjective trait level. Implicit measures are based on performance and they capture the mind in action. How quickly a respondent is able to make a quick association and a response is the rationale for the implicit association test (e.g., Steffens & Konig, 2006). If a respondent can more quickly classify a stimulus word such as conscientiousness to the "self" category compared to a "not self" category, then this person has a greater association for himself or herself being conscientiousness.

IAT scores can be altered deliberately, especially when respondents are given strategies to improve their scores (Gawronski, LeBel, & Peters, 2007). Fielder and Bluemke (2005) found that IAT scores were different when the participants were told to make their responses more favorable toward Turks than Germans. However, in most studies the amount of intentional distortion possible on an IAT is less compared to a self-report (e.g., Steffens, 2004).

In sum, each of the methods and measures for controlling has faking has advantages and disadvantages. Vasilopoulous and Cucina (2006) believe that methods for controlling faking may have introduced a cognitive aspect into the tests that should not be there. This is a problem because one of the reasons personality measures are in used in selection is because they do not exhibit adverse impact like cognitive ability measures (Vasilopoulous & Cucina, 2006). A personality measure that is not easily faked and does not induce a cognitive load should be used in selection.

Social Desirability Measures

Above I reviewed the methods and measures to detect and control for faking behavior. Researchers created these techniques because the first method to detect faking, a social desirability measure, is not perfect. The rest of this introduction reviews the research on social desirability measures and introduces the current study, testing the validity of a social desirability measure in an applied sample.

The first conceptualization of social desirability was as a response style, not a substantive trait (Burns & Christensen, 2006). The social desirability measures were created before the construct was defined (Burns & Christensen, 2006). The first “lie detector” scale was created by Ruch in 1942 and it was later used as the prototype of the K scale of the Minnesota Multiphasic Personality Inventory (MMPI; Ruch & Ruch, 1967). In 1957, Edwards created a social desirability scale as a measure of the tendency to give socially desirable responses in self-descriptions. Edwards (1957) found that social desirability scales correlated with personality measures if that trait was desirable (MMPI Neuroticism Scale correlates $r = -.50$ with Edward’s social desirability scale).

Developed in the 1940s, the MMPI is the most widely used clinical personality inventory; some of the scales assess individual emotional adjustment and attitudes toward test taking (Mesmer-Magnus & Viswesvaran, 2006). The MMPI has eight validity scales embedded within it: Cannot Say, Variable Response Inconsistency, True Response Inconsistency, Infrequency (F-scale), Fb (F-Back), Lie (L-scale), Correction (K-scale), and Superlative Self-Presentation Scale (S-scale). The L-scale scale assesses if a respondent is trying to present an overly positive image of him or herself, while the K-scale assesses if the respondent is attempting less blatant forms of faking. Various

combinations of scores on the validity scales require different interpretations (Mesmer-Magnus & Viswesvaran, 2006).

The California Psychological Inventory (CPI) is another widely used measure of personality for the normal adult population used particularly for personnel selection (Mesmer-Magnus & Viswesvaran, 2006). The CPI has three validity scales embedded within it: the Good Impression scale, the Communality scale, and the Sense of Well-Being scale. The Good Impression scale identifies respondents making a favorable impression, or “faking-good”. The Communality scale assesses random responding and the Sense of Well-Being scale assesses “faking-bad” whereas low scores indicate poor personal well-being. Researchers have shown that these scales are not effective at detecting a respondent trying to fake a certain trait (Mesmer-Magnus & Viswesvaran, 2006).

In 1960, researchers created the Marlowe-Crowne social desirability scale to address the limitations in Edward’s scale (see Zickar & Gibby, 2006). The full version is made up of 33 true or false items. The Marlowe-Crowne scale assumes that social desirability is a single, latent construct (Mesmer-Magnus & Viswesvaran, 2006).

In 1984, Paulhus found that social desirability scales do not correlate well with each other. Marlowe-Crowne and Edwards’ social desirability scale have a small relationship ($r = .24$). Paulhus (1984) examined the scales’ items loading on two factors, gamma and alpha, and Paulhus created his social desirability scale, the Balanced Inventory of Desirable Responding (BIDR), based on these two factors. Gamma is the conscious aspect of social desirability, called impression management, whereas alpha is

the unconscious aspect of social desirability, termed self-deceptive enhancement (Paulhus, 1984).

The BIDR is made up of two 20-item subscales: Impression Management (IM) and Self-Deceptive Enhancement (SDE). The scales are on a 5 or 7 point scale ranging from “not true” to “very true”. A sample item from the IM scale is “I never cover up my mistakes.”, and a sample item from the SDE scale is “I always know why I like things.” The IM scale correlates with traditional measures of faking such as the MMPI Lie scale, while the SDE correlates with measures of coping such as Edward’s SD scale (Paulhus, 2003).

Paulhus (1984) found that IM scale can be faked; while, SDE scale cannot be faked as much. However, recent research found both the IM scale and the SDE scale of the BIDR (as well as the Marlowe-Crowne Scale) can be faked by students under instructions to fake the scales (Pauls & Crost, 2004). The study examined mean differences in a within-subjects design of instructions to fake good (IM scores $M = 5.37$, $SD = 1.04$; SDE score $M = 5.20$, $SD = 0.74$) or respond honestly (IM scores $M = 3.12$, $SD = 0.87$; SDE score $M = 4.03$, $SD = 0.59$) (Pauls & Crost, 2004).

If an applicant fakes a personality measure, then they probably would distort the social desirability scale as well. If researchers are concerned with applicants scoring high on social desirability scales, then researchers should also be concerned with applicants scoring too low. Corrections for faking based on social desirability scores are ineffective because of the inability of researchers to ascertain applicants’ intentions. It is ironic that social desirability scales face the same problem as personality measures, when social desirability scales were created to address this limitation.

Correcting Faking with Social Desirability Measures

Industrial/Organizational psychologists and test administrators can correct personality scores for faking based on social desirability scales embedded within the test (Goffin & Christensen, 2003). One way to correct this is to disregard the personality tests of applicants with a high social desirability score. Another option is to lower the personality scores of those who score high on a measure of social desirability. The last option is using a social desirability score to adjust the personality measure score with a special equation created by the test developer (Goffin & Christensen, 2003). There are also the options for retesting the applicants and for interpreting the results of the test cautiously (Reeder and Ryan, 2011).

Some researchers believe corrections for faking are frivolous. Making corrections based on SD scores does not increase the validity of the test (Christiansen, Goffin, Johnston, & Rothstein, 1994) or increase mean performance (Schmitt & Oswald, 2006). Using social desirability scores to correct personality scores fail to produce a score that approximates a respondent's honest score (Ellingson, Sackett, & Hough, 1999). Removing the effect of social desirability on the Big Five personality traits leaves the criterion-related validity of the personality almost unchanged (Ones, Viswesvaran, & Reiss, 1996; Barrick & Mount, 1996).

Other research has examined the selection decisions made after correcting for SD. Christiansen et al. (1994) found that depending on the selection ratios, correcting scores resulted in different hiring decisions compared to uncorrected scores. In another study, researchers found that making corrections for faking either by the correcting scores or eliminating those high in SD, caused harm in their hypothetical selection procedure with undergraduates (Stewart, Darnold, Zimmerman, Parks, & Dustin, 2010). In some cases

with varying selection ratios, they retained 75% of those who faked and in some cases eliminated 35% of those who did not (Stewart et al., 2010).

Corrections for faking based on SD scores assume that SD is a suppressor variable (Burns & Christiansen, 2006). In theory, SD masks the relationship between the personality trait and job performance because it is correlated with the predictor but not the criterion. Reeder and Ryan (2011) make the case that the classical suppression view of SD is not feasible because correlations between personality and performance and personality and SD are not large enough for suppression to have an effect. Thus, corrections based on SD are not likely to produce large increments in validity unless the predictor-criterion and predictor-suppressor relationships are strong (Reeder & Ryan, 2011).

More recent research has looked at SD as a moderator of the personality-performance relationship (Burns & Christensen, 2006). SD as a moderator is classified as a trait and not a response style. With high levels of SD respondents' trait scores will have lower validity than those with low SD. Researchers who view SD as a trait suggested that corrections for social desirability lower criterion-related validity though a reduction of the relevant trait variance that social desirability has with other personality traits (Mueller-Hanson et al., 2003). Holden (2008) found using a moderated multiple regression analysis that using social desirability scales underestimated the effect sizes for induced faking on validity.

Based on SD scores, corrections can be made, applicants can be eliminated, or nothing can be done. However, even if nothing is done, the interpretation of the personality scores may be biased by the presence of SD scores. Christiansen, Rozek, and

Burns (2010) had practitioners ($N = 160$) in personnel selection read a job description and the assessment profiles of two hypothetical candidates. The researchers wanted to test the effects of differing SD scores on hiring judgments. The results showed when the two hypothetical candidates differed on the amount of SD responding, the individual with the higher score was judged as less candid and sincere. Also, those candidates with higher SD scores were rated as less hireable. Another interesting finding was that high scores on the personality test were not an indicator that the individual was less candid. The researchers found that decision-makers in selection make subjective mental adjustments to personality scores based on high SD scores, and they tend to regress the personality scores toward the mean (Christiansen, Rozek, & Burns, 2010). Thus, reporting SD scores in an assessment should not be taken lightly, because just their presence does have an effect.

The use of social desirability scales to detect and correct for faking is controversial, yet they are found in many personality tests and are used often in personnel selection. Goffin and Christensen (2003) found 12 widely used personality tests that have social desirability within them. Goffin and Christensen (2003) mailed a survey to 67 I/O psychologists asking them if they use a response validity scale when they administer personality measures for selection purposes. Thirty-six psychologists responded of whom 56% indicated that they use a personality measure that includes a response validity scale.

Social desirability as a trait. Researchers' views of SD are mixed. Some suggest that SD is faking (Ones, Viwesvarvan, & Reiss, 1996), while others believe that SD is one component of faking (Furman, 1986; McFarland & Ryan, 2006; Snell, Sydell, & Lueke, 1999). Ones, Viswesvaran, and Reiss (1996) conducted a meta-analysis of the SD

literature and found that SD does not function as a mediator, suppressor, or a predictor in personality testing for employment selection. SD was related to individual differences in personality. Conscientiousness had a low correlation ($r = .20$) with SD. SD correlated $r = .37$ with emotional stability, which was the highest correlation of SD with the Big Five (Ones, et al., 1996). Mesmer-Magnus, Viswesvaran, Deshpande, and Joseph (2006) found that a measure of social desirability, the Marlowe-Crowe, had a significant positive correlation with both self-esteem ($r = .20$) and emotional intelligence ($r = .44$).

Researchers examined the threat of SD to construct validity of personality measures. The data show that SD does not alter construct validity of the personality measures (e.g., Sisco & Reilly, 2010). In a study of four distinct groups of participants where some were not motivated to fake found SD overlaps with personality traits and does not represent distortion (Ellingson, Smith, & Sackett, 2001). Using factor analysis, they found that correlation matrices among the personality measures for high SD groups were similar to the correlation matrices for low SD groups. SD did not alter the construct validity of personality measures. Smith & Ellingson (2002) found the same patterns of results between job applicants and students that had no motivation to fake. They concluded that SD overlaps with personality traits and does not represent response distortion (Smith & Ellingson, 2002).

Holden and Passey (2010) tested if many different kinds of SD scales (BIDR-IM, BIDR-SDE, Marlowe-Crowne, Jackson's personality research form, and the validity index of the Hogan Personality Inventory) moderated the self-report Big Five personality measure and the peer-reported (the participants' roommates) Big Five personality measure. This was a low-stakes testing situation, i.e., no motivation to fake, the social

desirability measures had some substantial correlations with the self-report personality, but they correlated minimally with the peer-reported ratings. The researchers concluded that SD is a general self-report method variance that is unrelated to faking behavior. The BIDR impression management scale correlated $r = .48$ with agreeableness and $r = .28$ with conscientiousness (Holden & Passey, 2010).

Researchers who view SD as a trait usually use SD scores as a moderator of test validity and find that SD does not impact criterion-related validity or construct-related validity. These are considered traited SD studies, classified under the substance theory. In contrast, the situational SD studies, find that SD matters when comparing applicant to non-applicants. Researchers, who ascribe to the response style view of SD, usually find that SD decreases validity of personality measures and sometimes destroys the factor structure of personality measures.

In an effort to make sense of the SD literature, Henry and Raju (2006) examined the item-level and scale-level responses of a conscientiousness measure between groups classified as high and low on impression management (IM). Researchers examined the measurement equivalence of a personality measure given for selection versus a developmental administration (a situational view of SD) and for applicants with high IM versus applicants with low IM (a traited view of SD). They found minimal differential test and item functioning across both situational and traited IM investigations. The underlying meaning of the conscientiousness items remained intact across the group types. Thus, there are similarities between these two ways of looking at SD, as either a trait or a response style. Mean difference increases between groups in either case still mean that the scale is interpretable for both groups (Henry & Raju, 2006).

Holden and Book (2011) summarized and compared the research on applicant faking from laboratory studies and from studies in real-world settings. They looked at the effects applicant faking on mean scores, correlation structure, and validity. Laboratory studies usually find a medium effect size on personality scale mean scores differences between those instructed to respond honestly or fake. These effect sizes are typically larger than what is seen in studies involving real-world applicants. For the effects on correlational structure, faking studies conducted in the laboratory usually find that faking increases the convergence of personality scales, i.e., they become highly correlated, but this effect is not found in real-world studies, where subgroups of fakers and non-fakers are defined by a social desirability measure. Lastly for the studies that investigated the effects of faking on criterion-related validity, faking studies conducted in the laboratory usually find that faking reduces validity and the results for faking studies done with real-world samples are mixed. Sometimes faking does not affect validity of the personality measure (e.g., Barrick & Mount, 1996) and sometimes it does (e.g., Hough et al., 1990).

Holden and Book (2011) cite some discrepancies in the real-world studies that could account for the mixed results: differences in assessment contexts, the type of performance criteria, operationalization of faking (as either social desirability measures or applicants versus incumbents), analytic methods, or low-stakes versus high-stakes testing. Because researchers have shown that faking is a complex behavior that is determined by the situation and the person, it is no surprise that different types of sample may yield different results. The focus for my research is the operationalization of faking as something a social desirability score could capture. Based on previous research, faking classified this way typically shows no effect on correlational structure of personality

variables and criterion-related validity. To claim that faking does not have any ill effects on personality measures used in selection and place it on the back of a measure with limitations is imprudent.

Social desirability and job performance. If SD is conceptualized as a positive trait, then it should correlate well with job performance. However, some studies have found that this is not the case. Neither impression management (IM) nor self-deceptive enhancement (SDE) of the BIDR predicted performance strongly (Li & Bagger, 2006). The scales were weakly correlated to the performance measures IM ($\rho = .12$) and SDE ($\rho = .10$).

A meta-analysis found SD scores to have a mean correlation of $r = .04$ with managerial job performance (Viswesvaran, Ones, & Hough, 1999) Another meta-analysis found SD to be a predictor of training performance with mean correlation of $r = .19$ (Ones et al., 1996). However, these samples contained mostly incumbents. More recently, Viswesvaran, and Hough (2001) conducted a meta-analysis and found that impression management scores from a social desirability measure do not predict successful managerial job performance.

Researchers found that IM scores do not have a relationship with job performance and the validities of the personality measure remain intact regardless of high scores on IM (e.g., Ones et al., 1996; Li & Bagger, 2006). Berry, Page, and Sackett (2007) explored whether this held true for Self-Deceptive Enhancement (SDE), the second subscale of Paulhus' Balanced Inventory of Desirable Responding, a social desirability measure. In their study with 277 managers at a large energy company, they found that accounting for main effects of SDE scores and the interaction between SDE and each of the Big Five

personality traits increased the prediction of job performance in a hierarchical regression analysis, but the same effect was not found for the IM scale. The effect of increased job performance prediction when accounting for SDE occurred because SDE is a suppressor of the emotional stability and job performance relationship for this sample of managers (Berry, Page, & Sackett, 2007).

Burns and Christensen (2006) call for more studies linking SD to job performance. In addition, using an applied sample of applicants instead of college students in a laboratory will be useful because it is more realistic and reflective of what occurs in personnel selection.

Literature Review Conclusions

The evidence suggests that faking on personality measures is possible and occurs within selection. The biggest concern for employers are applicants with low levels of the desired trait that fake to move upwards in the selection process or eventually get hired. There are many methods and measures to control and detect applicant faking, yet none is the panacea. The oldest and most common method of detecting faking, social desirability measures, continue to be used despite concrete evidence that they are effective proxies of applicant faking.

In their review of the social desirability research, Burns and Christensen (2006) highlight some of the unanswered questions about social desirability measures and call for more research to be conducted. Some of the “unanswered” questions involve the relationship between social desirability scores and job performance, whether social desirability assesses a single construct or a composite of personality traits, and if social desirability scores reflect self-deception or impression management (Burns &

Christensen, 2006). Despite the large amount of research on social desirability measures and concerns about their effectiveness, they still are used in selection as applicant faking detectors.

The Current Studies: Purpose and Contributions to the Literature

The purpose of this research is to address and fill gaps in the literature on faking and social desirability measures. Griffith and Peterson (2008) stated that for SD measures to be useful in applied setting and in research empirical evidence must test the assumption that these measures are reliable and valid. Even just the presence of SD scores encourages decision-makers in selection to make subjective decisions about the personality scores (Christiansen, Rozek, & Burns, 2010). I am interested in effectiveness and usefulness of having SD measures in a personnel selection assessment.

Most researchers have concluded that social desirability measures are not effective indicators of faking behavior (e.g., MacCann, Ziegler, Roberts, 2011); however, they are still frequently used in selection as response validity scales (Goffin & Christensen, 2003). I am is interested in the implications of having social desirability scales in selection; in particular, if SD scales are able to identify fakers who increase their score from one administration of a test to the next and examining how cutoffs and corrections for faking based on SD scores affect job performance at varying selection ratios. I am concerned for organizations making selection decisions based in part on SD measures. I will extend the current research on SD measures to see if they are able to detect faking behavior in an applied sample.

Faking behavior is characterized as a complex interaction of person, test, and the situation (e.g., McFarland & Ryan, 2006). Someone who takes a personality test and does

not get the job and takes the test again, will likely change their score to increase the likelihood of getting hired. Landers, Sackett, & Tuzinski (2010) suggested that those who are retesting after failure on a personality test and have low initial test scores are more likely to respond differently and improve their scores. Thus, the test-retest reliability of a personality measure will be lowered when an applicant takes the same measure twice. The amount of faking as measured with a social desirability measure may be different between two administrations as well. Researchers found that retest effects varied by subgroup, such that females and younger candidates improved more upon retesting than did males and older candidates (Van Iddenkige, Morgeson, Schleicher, & Campion, 2011).

One of the purposes of SD measures is to detect faking behavior. However, many researchers doubt their effectiveness and their ability to identify fakers (e.g. Griffith & Peterson, 2008). Hypotheses 1 and 2 were tested with an archival dataset of participants who took the same personality measures on two separate occasions. The personality measures used for Study One are described below in the measures section of this document. The assessment also included a scale similar to Paulhus' (1998) Impression Management scale of the BIDR, a social desirability measure. The impression management scale is used to detect those who are managing their impressions.

Hypothesis 1: Participants are able change their personality scores on different administrations of a test.

To detect response distortion and test the efficacy of the impression management scale, I applied a method to identify fakers used by Arthur Jr., Glaze, Villado, and Taylor, (2010). At both administrations, the participants were either applicants/applicants,

incumbents/applicants, or applicants/incumbents. The participants who were applicants on both administrations of the test were used as the baseline group to detect the average amount of score change possible on the personality measures. The formula of the standard error of measurement difference score (SEM_d) was computed on the baseline group. For the individuals who were incumbents/applicants or applicants/incumbents, incumbent scores were subtracted from applicant scores. If this number exceeded the SEM_d computed on the baseline group, he or she was classified as a faker on that personality measure.

A recent study found that those flagged by a social desirability measure had higher personality scores than those who were not, regardless if they were applicants or incumbents (O'Connell, Kung, & Tristan, 2011). The standardized mean difference between flagged faker and non-faker was similar for the applicant sample ($d = 2.69$) and the incumbent sample ($d = 2.49$). The O'Connell et al. (2011) study used a between-subjects design; I extend their work by using a within-subjects design which has higher statistical power. Within-subjects designs also avoid the confounding effect of between-subjects variance. With within-subjects designs it is also possible to evaluate individual level changes in faking. Ellingson et al. (2007) utilized a within-subjects design, but their study had several limitations such as, an effect for time, feedback, and the use of personality scales that are more generic and not work related.

Hypothesis 2: Impression Management scales are able to identify fakers of personality measures.

Hypotheses 1 and 2 make up Study One and they examine whether social desirability (impression management) scales are able to detect of faking behavior. If they

are unable, there is further evidence not to use them in selection. Hypotheses 3 - 5 are the first step to test if using impression management measures to detect and correct for faking harms the organizational performance. Research Questions 1 - 4 will illustrate the effects of impression management measures on job performance. I hypothesized that impression management is a necessary skill for salespeople, thus selecting people out or correcting their personality scores because of it results in worse people hired and lowers organizational performance.

Like Study One, Study Two uses the personality measures of extraversion and competitiveness. I hypothesize that these two measures each will have a relationship with job performance in this sample of salespeople. Extraversion has been linked to job performance in past studies (e.g., Barrick, Mount, & Judge, 2001). Salespeople need to be able to prospect and doing so requires them to effectively network and communicate with strangers. Salespeople are also often placed in head-to-head combat against each other to achieve sales targets for the organization. They compete against other companies to land and maintain an account, thus an aggressive, competitive drive is needed to succeed.

Hypothesis 3a: There is a positive relationship between extraversion and weekly sales revenue.

Hypothesis 3b: There is a positive relationship between competitiveness and weekly sales revenue.

Hypothesis 4a: There is a positive relationship between extraversion and supervisor's ratings of job performance.

Hypothesis 4b: There is a positive relationship between competitiveness and supervisor's ratings of job performance.

To my knowledge, there have been no studies that examine SD scores from applicants to their job performance in the role, but researchers have called for this research to be done. Burns and Christensen (2006) called for more studies linking SD to job performance and to use applicant samples focusing on specific jobs. Viswesvaran, Ones, and Hough (2001) declared that future research should examine the relationship between impression management and job performance for sales jobs.

For Study Two, I examined the SD and job performance relationship for a sales job. Researchers who believe that faking is determined by a social desirability scale find that faking does not have a relationship with job performance (Ones, Viswesvaran, & Reiss, 1996; Viswesvaran & Hough 2001). However, researchers believe that there may be some instances where successful faking on personality measures may be a good thing for certain jobs, e.g. sales jobs (Hogan, Hogan, & Roberts, 1996). Part of salespeople's success depends on them being able to tailor their image to various stakeholders. They interact with many types of people daily and need to place themselves in a favorable light to make the sale. For this research, I used a scale similar to Paulhus' (1998) Impression Management scale of the BIDR, a social desirability measure. The impression management scale is used to detect those who are managing their impressions. Thus, sales job applicants' impression management scores should have a positive relationship to their job performance as incumbents.

Only a few studies have examined the social desirability and job performance relationship and used other measures of performance instead of overall performance, such as, task performance, contextual performance (e.g., O'Connell, Kung, & Tristan, 2011).

My study is unique that it measures two types of performance: supervisor ratings and weekly sales generated revenue.

Hypothesis 5a: Salesperson's impression management is positively related to their weekly sales revenue.

Hypothesis 5b: Salesperson's impression management is positively related to their supervisor's rating of job performance.

Because impression management has a positive relationship with job performance for salespeople, selecting out individuals who are faking based on their IM scores, would lower the job performance of the sample hired. I conducted a simulation with the dataset, varying the proportions removed for suspected faking based on their IM score (the top 6% removed and also the top 24% removed). I looked at these differences across several selection ratios. Research Questions 1a and 1b investigated whether removing salespeople based on their high IM scores results in lower potential sales dollars and lower supervisor ratings for the organization compared to the sample with no one removed. Research Question 2 compared this effect between weekly sales revenue and supervisor ratings.

Research Question 1a: Does selecting out salespeople based on high impression management scores result in lower weekly sales revenue for the organization?

Research Question 1b: Does selecting out salespeople based on high impression management scores result in lower supervisor ratings?

Research Question 2: Does selecting out salespeople based on high impression management scores have a larger effect on weekly sales revenue than supervisor ratings?

Corrections are made to personality scores based on IM scores to estimate an honest score for a respondent. Researchers have called for studies that examine organizational outcomes based on making corrections to personality scores (Reeder & Ryan, 2011). The personality measures of extraversion and competitiveness were corrected based on the participants' impression management scores. If the participant scored above the 76th percentile on the impression management and had an extraversion score above the average for the group his or her extraversion scores was changed to the average extraversion score. If the participant scored above the 76th percentile on the impression management and had a competitiveness score above the average for the group his or her competitiveness scores was changed to the average competitiveness score.

Since impression management should have a positive relationship with salesperson's job performance, correcting personality scores in selection results in worse people being hired. If an applicant scores high on IM, their personality scores will be corrected (reduced) and he or she is less likely to be selected. I examined this relationship at varying selection ratios for both average weekly sales revenue and supervisor ratings.

Similar to Research Questions 1 and 2, the following Research questions investigated how corrections impacted job performance.

Research Question 3: Does correcting salespeople's personality scores on the basis of impression management result in lower weekly sales revenue for the organization?

Research Question 4: Does correcting salespeople's personality scores on the basis of impression management result in lower supervisor ratings for the organization?

Research Question 5: Does making corrections to personality scores based on high impression management scores have a larger effect on weekly sales revenue than supervisor ratings for the organization?

II. STUDY ONE

Method

Participants

The archival data from Study One participants came from a Midwest private management and salesperson consulting firm which carries out web-based assessments for multiple client organizations for the purposes of selection and employee development. There were 3,471 participants in Study One; these participants completed the assessment twice and it was possible to match their data from both administrations. This sample only included participants with 60 months or less between testing administrations ($M = 24.01$ months; $SD = 16.22$). Participant data were matched by last name, first name, middle initial, gender, and race. Sixty-five percent of the participants are male and 73% are Caucasian, 12.6% are African-American. The average age was 35 years old ($M = 35.26$; $SD = 9.5$). Of the total sample, 3,136 participants were part of a validation study (applicants) at time one and time two, 206 participants were part of a validation study (incumbents) at time one and applicants at time two, and 129 participants were applicants at time one and incumbents at time two. The demographic composition of the three groups was very similar.

Most of the sample ($N = 3,471$) was assessed for sales positions, 65.58% at administration one and 60.04% at administration two. Fourteen percent (14.26%) were assessed for managerial positions at administration one, and 13.74% at administration

two. Twelve percent (12.01%) were assessed at administration one for administration /professional staff/technical positions, and 11.99% at time two. Only a small portion was assessed for developmental reasons, 1.12% at time one and 0.63% at time two. In the archival dataset, 10.02% did not have a position type listed for time one and 13.60% did not have a position type listed for time two. Regardless of which position a participant was being assessed for, they received the same assessment.

Measures

Extraversion. Items for the extraversion scale were taken from a proprietary assessment instrument designed to measure relevant work behaviors and personality constructs for on-the-job effectiveness. The assessment is made up of empirically-keyed scales. One of the scales within this assessment measures effective networking, and this scale has demonstrated convergent validity with Goldberg's International Personality Item Pool (IPIP, 2001) measure of extraversion ($r_c = .77$). The scale is made up of 24 statements. Respondents are required to answer either "True" or "False" to the statements. An example item is: "It does not bother me to give presentations to groups." The scale alpha is .70.

Competitiveness. Items for the competitiveness scale were taken from the same proprietary assessment instrument described above. One of the scales within this assessment measures competitiveness, which is described as someone who places self and others in head-to-head competition and is unwilling to relent or give up. The scale is made up of 13 statements. Respondents are required to answer either "True" or "False" to the statements. An example item is: "Only the fittest survive." The scale alpha is .61.

Impression Management. One scale in the proprietary assessment is used to measure if the individual is managing his or her impressions, which is similar to Paulhus's (1998) Impression Management scale. It was developed based on MMPI's Lie Scale (Butcher et al., 1992). The scale is made up of nine statements. Respondents are required to answer either "True" or "False" to the statements. An example item is: "Criticism from others never upsets me." The scale alpha is .70. The proprietary scale was reverse scored for this study, so that a high score corresponds to more of the construct.

Procedure

The participants completed a proprietary assessment that was designed to measure relevant work behaviors and personality constructs for their job (see Miller et al., 2005 for a full description). The dataset used for all the following analyses for Study One included only participants who completed the proprietary assessment twice, and it was possible to match both administrations.

Results

Tests of Hypotheses

Hypothesis 1. The results for Hypothesis 1 are displayed in Table 1. The results supported Hypothesis 1, as there were statistically significant within-subject mean differences on the personality scores between the two administrations, but only if participant's status as applicant or incumbent changed on the next administration. Paired samples t-tests were calculated to assess if there were mean scores changed across the two administrations. As mentioned in the Participants section of this document, groups were constructed within the archival dataset based if the participant was an applicant/applicant, incumbent/applicant, or applicant/incumbent on the two administrations.

Group One participants were applicants on both administrations of the test and their extraversion mean scores did not differ between time one ($M = 18.30$; $SD = 3.16$) and time two ($M = 18.28$; $SD = 3.19$, $t(3135) = 0.39$, $p = .70$). Statistically different differences were not found for the competitiveness scores for Group One, time one ($M = 4.71$; $SD = 2.06$) and time two ($M = 4.67$, $SD = 2.08$, $t(3135) = 1.03$, $p = .30$).

For the extraversion measure, applicant scores were higher than incumbent scores. Group Two participants were incumbents at time one ($M = 17.88$; $SD = 3.06$) and applicants at time two ($M = 18.52$; $SD = 2.78$); there were significant mean differences between administrations ($t(205) = -2.57$, $p < .05$). Group Three participants were applicants at time one ($M = 18.33$; $SD = 3.08$) and incumbents at time two ($M = 17.49$; $SD = 3.59$); there were significant mean differences between administrations ($t(128) =$

2.40, $p < .05$). The Cohen's d analyses revealed a small mean effect for both Group Two ($d = -0.22$) and Group Three ($d = 0.25$).

For the measure of competitiveness, incumbent scores were higher than applicants. Group Two participants were incumbents at time one ($M = 5.28$; $SD = 2.39$) and applicants at time two ($M = 4.61$; $SD = 1.98$); there were significant mean differences between administrations on the competitiveness measure ($t(205) = 3.54$, $p < .01$). Group Three participants were applicants at time one ($M = 4.53$; $SD = 2.25$) and incumbents at time two ($M = 5.14$; $SD = 2.17$); there were significant mean differences on the competitiveness scores between administrations ($t(128) = -2.47$, $p < .05$). The Cohen's d analyses revealed a small mean effect for both Group Two ($d = 0.31$) and Group Three ($d = -0.28$).

The reason for these lower applicant scores for the competitiveness measure could be the social desirability of the scale. Typically applicants raise their scores on personality measures to be seen in a more favorable light. Participants may have been reticent to endorse item options that would load higher on competitiveness as they may have feared being perceived as cut-throat, self-interested, and callous. Applicants may have wanted to appear more agreeable and thus lowered their score on the measure. Figures 3 and 4 show how the majority of participants' scores are at the lower end of the scale.

Hypothesis 2. The second hypothesis stated that the impression management scale is able to identify fakers of personality measures. To test this hypothesis, a faking indicator was created by establishing a baseline for how much scores typically change. This was created on Group One, i.e., the participants who were applicants for both for

administrations of the assessment. The formula of the standard error of measurement of the difference scores (SEM_d) was computed on Group One,

$$SEM_d = S_d \sqrt{1 - r_{x1x2}} \quad (1)$$

where S_d is the standard deviation of the time 2 – time 1 difference score and r_{x1x2} is the time 1/time 2 correlation (Arthur Jr., et al., 2010). For extraversion ($M_d = -.02$; $S_d = 3.34$; $r_{x1x2} = .45$) the SEM_d was 2.48. For competitiveness ($M_d = -.04$; $S_d = 2.34$; $r_{x1x2} = .37$) the SEM_d was 1.85.

Groups Two and Three's incumbent scores and applicant scores on the measures were combined into two metrics. Applicant scores for Group Two came from the second administration of the test and applicant scores for Group Three came from the first administration of the test, and vice versa for incumbent scores. To determine if there were order effects that would contaminate the incumbent and applicant scores, independent-samples t-tests were conducted. There was no significant difference between incumbent extraversion scores from time one (Group Two $M = 17.88$; $SD = 3.06$) and incumbent extraversion scores from time two (Group Three $M = 17.49$; $SD = 3.59$); $t(333) = 1.04$, $p = 0.28$. There was no significant difference between applicant extraversion scores from time one (Group Three $M = 18.33$; $SD = 3.08$) and applicant extraversion scores from time two (Group Two $M = 18.52$; $SD = 2.78$); $t(333) = 0.57$, $p = 0.57$.

For competitiveness, there was no significant difference between incumbent competitiveness scores from time one (Group Two $M = 5.28$; $SD = 2.40$) and incumbent competitiveness scores from time two (Group Three $M = 5.14$; $SD = 2.17$); $t(328) = 1.01$, $p = 0.31$. Finally, there was no significant difference between applicant competitiveness scores from time one (Group Three $M = 4.53$; $SD = 2.25$) and applicant

competitiveness scores from time two (Group Two $M = 4.61$; $SD = 1.99$); $t(328) = 0.36$, $p = 0.72$.

For impression management scores, there was no significant difference between incumbent impression management scores from time one (Group Two $M = 5.87$; $SD = 1.79$) and incumbent impression management scores from time two (Group Three $M = 5.97$; $SD = 1.74$); $t(332) = -0.51$, $p = 0.61$. Finally, there was no significant difference between applicant impression management scores from time one (Group Three $M = 6.36$, $SD = 1.62$) and applicant impression management scores from time two (Group Two; $M = 6.21$, $SD = 1.77$); $t(332) = -0.76$, $p = 0.45$.

The faking indicator (SEM_d) was used to identify fakers. Computed from Group One, the SEM_d for extraversion was 2.48 and the SEM_d for competitiveness was 1.85. Since statistically significant order effects did not appear for incumbent and applicant personality scores, Groups Two and Three were combined. In this combination group, each participant's incumbent personality score was subtracted from his or her applicant score. If that number exceeded the SEM_d for that personality measure, he or she was identified as a faker. Of the sample of 335, 24.8% were identified as faking the extraversion scale and 20.6% were identified as faking the competitiveness scale.

Results supported Hypothesis 2 as fakers were identified by the impression management scale. The results are displayed in Table 5. The incumbent impression management scale correlated with the extraversion faking indicator negatively ($r = -.21$, $p < .01$). The applicant impression management scale correlated with the extraversion faking indicator positively ($r = .12$, $p < .05$). The incumbent impression management scale did not have a relationship with the competitiveness faking indicator ($r = .09$, $p =$

.10). The applicant impression management scale correlated with the competitiveness faking indicator negatively ($r = -.22, p < .01$).

The within-subjects mean differences are displayed in Table 6. Extraversion scores from when the participants were incumbents are *lower* for those identified as fakers ($M = 13.88; SD = 5.27$) than those identified as non-fakers ($M = 19.00; SD = 2.13$). The opposite effect occurred with the extraversion applicant scores; non-faker means were *lower* ($M = 18.21; SD = 3.14$) than faker ($M = 19.16; SD = 1.83$).

Similar to the effect above, competitiveness scores from when the participants were incumbents are *lower* for those identified as fakers ($M = 3.36; SD = 1.70$) than those identified as non-fakers ($M = 5.71; SD = 2.20$). The opposite effect occurred with the competitiveness applicant scores; non-faker means were *lower* ($M = 4.06; SD = 1.74$) than faker ($M = 6.55; SD = 2.17$).

The explanation for the applicant scores is intuitive. Those identified by the faking indicator, who raised their scores beyond the SEM_d , had higher applicant personality scores than incumbent personality scores. However, for the non-fakers' scores, incumbent scores were higher than their applicant. The non-fakers' incumbent personality scores were already at the high end of the scale, so they had limited opportunity to raise their score and be identified. This ceiling effect is shown for the extraversion measure in Figure 2, as most score at the high end of the scale. The floor effect is shown also in Figure 4 for competitiveness; more are scoring at the low end of the scale.

Hypothesis 2 was concerned with how well impression management scales were able to identify faking behavior. As an exploratory analysis, I conducted a discriminant

analysis on participants classified as fakers or non-fakers and their impression management scores from incumbent and applicant administrations. A simple way to explain discriminant analysis is that it is reverse regression. The independent variable of impression management incumbent scores and impression management applicant scores are used to predict fakers or non-fakers. The analyses were conducted with prior probabilities selected. Classification results indicate how well membership can be predicted for each group (faker or non-faker) for impression management. Results are displayed in Tables 7 and 8. For participants identified as faking extraversion, the percentage of correctly classified was 75.4%. The percentage correctly classified as faking competitiveness or not was higher, at 81.1%. These percentages are relatively high.

III. STUDY TWO

Method

Participants

Data from the Study Two participants comes from a Midwest private management and salesperson consulting firm which carries out web-based assessments for multiple client organizations for the purposes of selection and employee development. Study Two participants comprise a sample of sales account executives ($N = 382$) who completed the proprietary assessment as applicants and then were hired by the Midwest logistics sales firm. They had been in the job for at least 12 weeks ($M = 65.38$; $SD = 36.34$). The average age was 26 years old ($M = 26.08$; $SD = 5.52$), 91.4% of the sample is white, and 83.4% are male.

Measures

The measures of Extraversion, Competitiveness, and Impression Management are used in Study Two. They are described above in the Measures section of Study One.

Sales Revenue. Job performance for the salespeople was weekly brokerage revenue, or the sales dollars the sales account executive generated that week. This was averaged to create an average dollar variable.

Supervisor Ratings. Each participant was rated by four supervisors. This was averaged to create an average supervisor rating. The ratings were on a three point scale: 1 “Does not meet expectations”, 2 “Meets Expectations”, and 3 “Exceeds Expectations”.

Procedure

In 2007 through 2008, the sales account executives completed a proprietary assessment which was designed to measure relevant work behaviors and personality constructs for their job (see Miller et al., 2005 for a full description). In 2010, the consulting firm received performance metrics for the sales account executives. The assessment data were matched to the job performance metrics.

Results

Tests of Hypotheses

The hypotheses and research questions for Study Two were conducted on an archival dataset of 382 salespeople. These salespeople responded to measures of extraversion, competitiveness, and impression management as applicants. The archival dataset contains the salesperson assessment data as well as their job performance data (weekly sales revenue and supervisor ratings). They had been in the role for a minimum of 12 weeks.

Hypotheses 3 - 5 and Research Questions 1 - 4 examined if impression management scores used to detect and correct for faking harms organizational performance. I hypothesized that impression management is a necessary skill for salespeople, thus selecting people out or correcting personality scores because of impression management results in worse people hired and lowers organizational performance. Table 9 displays the results.

Hypothesis 3a. Hypothesis 3a was not supported. Extraversion does not have a relationship with weekly brokerage dollars ($r = -.04, p = .34$).

Hypothesis 3b. Hypothesis 3b was supported. Competitiveness does have a relationship with weekly brokerage dollars ($r = .16, p < .01$).

Hypothesis 4a. Hypothesis 4a was not supported. Extraversion does not have a relationship with supervisor ratings ($r = -.06, p = .63$).

Hypothesis 4b. Hypothesis 4b was supported. Competitiveness does have a relationship with supervisor ratings ($r = .13, p < .01$).

Only competitiveness, not extraversion, had a relationship with job performance in this sample. The correlation coefficient was small, but similar to other validity coefficients for personality measures (Barrick, Mount, & Judge, 2001). Since competitiveness has a relationship with job performance it would make sense only to look at this measure for the effects of eliminating or correcting competitiveness scores because of impression management. However, the results of Hypothesis 5 show that impression management does not have a relationship with job performance.

Hypothesis 5a. Hypothesis 5a was not supported. Impression management does not have a relationship with average brokerage dollars ($r = -.07, p = .45$).

Hypothesis 5b. Hypothesis 5b was not supported. Impression management does not have a relationship with supervisor ratings ($r = -.04, p = .12$).

Hypothesis 5 was not supported as impression management did not have a statistically significant relationship with job performance. I hypothesized that it would and that eliminating applicants based on impression management would harm job performance. Research Questions 1 and 2 shown below further examine this relationship. Although, because there is no relationship between impression management and job performance for this sample of salespeople eliminating individuals based on impression management scores would not harm organizational job performance.

Research Questions

The research questions examine how eliminating participants from the selection pool or correcting their personality scores both based on impression management scores affects job performance. This was examined at varying selection ratios: low, medium,

and high. In order to make clean splits based on scores, the selection ratio had to be determined by the percentage of people with a certain personality score. For example, the lowest selection ratio of .32 was selected because 32% of the sample had the highest extraversion score of 21. The medium selection ratio of .55 was selected because 55% of the sample had extraversion scores of 21 or 20. The highest selection ratio of .72 was selected because 72% of the sample had scores of 21, 20, or 19. For competitiveness, the lowest selection ratio .11 had competitiveness scores of 12, 11, 10, or 9. The selection ratio of .28 had competitiveness scores of 12, 11, 10, 9, 8, or 7. Finally, the selection ratio of .64 had the competitiveness scores of 12, 11, 10, 9, 8, 7, 6 or 5.

Research Question 1a: Does selecting out salespeople based on high impression management scores result in lower weekly sales revenue for the organization?

Selecting out salespeople based on high impression management scores does not result in lower weekly sales revenue. Table 10 displays the results for extraversion. Extraversion does not have a relationship with job performance (Hypotheses 3a and 4a). Higher weekly sales revenue at lower selection ratios was expected. This pattern does not appear; job performance increases as the selection ratio increases. The results for Hypothesis 5a showed that impression management does not have a relationship with weekly sales dollars. Therefore, the effect of decreased validity was not shown. Removing participants from the selection simulation because they scored at the top 7% of the impression management scale or the top 24% does not harm job performance. Removing participants because of impression management increases weekly sales revenue.

Table 12 displays the results for competitiveness. Competitiveness does have a relationship with job performance (Hypothesis 3b and 4b). Weekly sales revenue decreased as the selection ratio increased. Similar to the effect for extraversion, removing participants from the selection pool based on their impression management score increased job performance not harmed it.

Research Question 1b: Does selecting out salespeople based on high impression management scores result in lower supervisor ratings?

Research Question 2: Does selecting out salespeople based on high impression management scores have a larger effect on weekly sales revenue than supervisor ratings?

The effects for supervisor ratings are shown in Tables 11 and 13. Like above for extraversion, ratings increased as the selection ratio increased. For competitiveness, ratings decreased as the selection ratio increased. Average supervisor ratings were similar between the groups. There was not much change because of removing participants for their impression management scores. Selecting out salespeople based on high impression management scores does not result in lower supervisor ratings. To answer Research Question 2 is difficult because it is hard to compare change because the supervisor ratings changed so little between conditions.

Similar to Research Questions 1 and 2, the following research questions investigated how corrections for faking impacted job performance. If the participant scored above the 76th percentile on the impression management scale and had an extraversion score above the average for the group his or her extraversion scores was

changed to the average extraversion score. The average extraversion score was 19. Sixty-two of the 382 participants had their extraversion score corrected, a rate of 16.23%.

If the participant scored above the 76th percentile on the impression management scale and had a competitiveness score above the average for the group his or her competitiveness scores was changed to the average competitiveness score. The average competitiveness score was 5.41, and only 18 participants (of 381) had their score competitiveness score corrected, a rate of 4.72%.

Research Question 3: Does correcting salespeople's personality scores on the basis of impression management result in lower weekly sales revenue for the organization?

The results are mixed. Weekly sales revenue for corrected extraversion scores was lower than non-corrected at varying selection ratios, while corrected competitiveness scores had higher weekly sales revenue compared to non-corrected. Table 14 displays the results for extraversion. Like the results for Research Question 1, sales revenue increases as the selection ratio increases. When participant's extraversion scores were reduced to the average for high impression management, weekly sales dollars decreases. Making corrections hurt organizational performance.

This effect was not found for the competitiveness scores (see Table 16). Corrected competitiveness scores were higher than uncorrected. Fewer people had their scores corrected, only 18 participants for competitiveness, while 62 participants had their extraversion scores corrected to the mean. It is more important to look at the changes competitiveness scores; because competitiveness does have a relationship with job

performance, but so few scores were changed it is hard to make a meaningful comparison.

Research Question 4: Does correcting salespeople's personality scores on the basis of impression management result in lower supervisor ratings for the organization?

Competitiveness not extraversion has a relationship with supervisor ratings, so it logical to only look at those results. Supervisor ratings for the corrected competitiveness scores were very similar to the uncorrected, because so few participants had their scores corrected. Supervisor ratings did decrease and the selection ratio increased.

Research Question 5: Does making corrections to personality scores based on high impression management scores have a larger effect on average sales revenue than supervisor ratings for the organization?

Answering Research Question 5 is difficult. It is hard to compare change between the two types of job performance metrics because the supervisor ratings changed so little between conditions.

IV. DISCUSSION

Summary

My research explored the implications of having impression management (IM) scales in selection. In particular, if IM scales are able to identify fakers who increased their score from one administration of a test to the next. In Study Two, I examined how cutoffs and corrections for faking based on IM scores affect job performance at varying selection ratios. I postulated that impression management was a necessary trait for salespeople and that removing people or correcting their scores based on IM would harm job performance. IM did not have a relationship with job performance, thus that effect was not found.

At first glance it looks like impression management scales, are useful in selection. The results from Study One showed that they were able to identify fakers. Those who had high impression management scores as applicants were flagged as faking extraversion. For competitiveness, the relationship between applicant impression management scores and being identified as a faker was negative. On average, applicant competitiveness scores were lower than incumbent scores. Those responding as applicants were trying to make themselves look better by lowering their scores.

The faking indicator, the standard error of measurement difference score (SEM_d), was computed on Group One those who were applicants on both administrations. For the individuals who were incumbents/applicants or applicants/incumbents, incumbent scores were subtracted from applicant scores. If this number exceeded the SEM_d computed on the baseline group, he or she was classified as a faker on that personality measure. For competitiveness, participants in the applicant context lowered their scores, so the

difference of applicant - incumbent scores, is unlikely to exceed that threshold, thus he or she would not be classified as a faker. He or she was faking by lowering his or her scores to appear more desirable on the competitiveness measure.

A recent study found that those flagged by a social desirability measure had higher personality scores than those who were not, between the groups of applicants or incumbents (O'Connell, Kung, & Tristan, 2011). My study did not find these effects within-groups. My sample had scores from when the participants were incumbents and applicants. Applicants identified as fakers had higher mean scores than non-fakers for both personality measures, but incumbents identified as non-fakers had higher mean scores than fakers. Being identified as a faker with the faking indicator depends on your honest score. If the participant was already high enough on the scale there was no room for improvement. There was a ceiling effect (floor effect) for the incumbent personality measures as most people were scoring at the end (or beginning for competitiveness) of the scale. Thus, impression management scores are useful to identify faking but under certain conditions. Impression management only works to identify fakers if the person is low enough on the trait to have opportunity to improve their scores and if they are not trying to lower their scores to look better, like on competitiveness.

Models of faking behavior position faking as the interaction of dispositional factors experiential factors, and test characteristics that influence the ability to fake (Snell, Sydell, & Lueke, 1999). Opportunity to fake is one such characteristic; if one is already high enough on the construct, he or she has no opportunity to improve his or her score. This individual could either have a true high score on the measure or the items are transparent enough that the individual has knowledge of the construct they are trying to

measure. This opportunity to fake explains the ceiling effect. However, it is not possible to know if the person is truly high on the construct or was just able to raise his or her score on the scale.

The results from Study Two showed that removing those who score high on impression management increased the job performance of the sample. I expected the opposite effect. Extraversion did not have a relationship with job performance, so it is unproductive to examine its effects at varying selection ratios. Weekly sales revenue increased when those scoring high on IM were removed when salespeople were selected based on high competitiveness scores. The weekly sales dollar does not seem high, but when it is multiplied by 52 weeks in a year it can show a significant difference. For example, at the lowest selection ratio the full sample has yearly sales revenue of \$64,593.36, while if those scoring above the 76th percentile on impression management were removed the yearly sales revenue for that group would be \$65,799.76. If it was a sales force of ten people that difference would be \$120,640 yearly for the organization.

Supervisor ratings remained similar across groups. At the lowest selection ratio, the full sample with no one removed had average ratings of 2.10. Those scoring above the 76th percentile on impression management had average ratings of 2.12. It is hard to compare changes in dollars to changes in supervisor ratings, but it can be inferred that removing applicants who score high in impression management has more of an effect on objective measures of performance (dollars) than subjective measures of performance (ratings).

When looking at corrected personality scores versus not, weekly sales revenue was higher for the sample with competitiveness scores that were corrected because of high impression management. At the yearly level for the .28 selection ratio, the sample who was selected because of corrected competitiveness score had sales revenue of \$50,334.44. The sample with uncorrected scores had sales revenue of \$49,004.80. The yearly difference for an organization with ten salespeople would be \$132,964. No discernible pattern emerged for the supervisor ratings.

Implications

The possible implications for organizations are that if they want to increase job performance they should remove participants high in impression management or correct their personality scores. Sales revenue increased for the sample removed because of high IM scores and for the sample with corrected competitiveness scores based on high IM. It is important for organizations to examine the relationship between IM and job performance before making any selection decisions based on it.

The implications for academic research in Industrial/Organizational Psychology are that impression management measures are not true measures of faking for a couple of reasons. One, it cannot be determined from this study if IM scales are fokable themselves. Two, impression management from applicants correlated $r = .14$ with the extraversion faking indicator. If impression management was the perfect proxy for faking, we would expect to see a much high correlation. There are two camps of researchers those who think faking does not matter because SD does not affect validity (e.g., Ones et al., 1996) and those who think faking cannot be measured with a SD scale (e.g., Griffith & Peterson, 2008). I agree with the second camp that people are able to fake personality

scores, look at how the mean scores change between applicant and incumbent for this within-subjects sample. This study found that IM has a small relationship with faking, but the IM scale could only identify those who have the opportunity to fake. I predict that impression management scales used as response validity measures will be used less frequently, as more and more researchers are condemning their use and as innovative techniques to measure personality that are less fakable become more prevalent.

Limitations

There are several limitations of this study. There was poor test-retest reliability of the measures. Incumbent and applicant personality scores correlated $r = .27$ for extraversion and $r = .21$ for competitiveness. In addition, there is no way to know if the people who took the test again, what their situation was. If they were applicants, we do not know if they were currently employed by another company or if they were in their current company just testing for another position. It also could have been the case where their managers wanted them to retake the test to get a better score. If they had a better score, they would be more promotable.

Another limitation was the distribution of the personality measures. It was not a bell-shaped curve as most fell towards the end (or the beginning for competitiveness) of the scale. There was less variability in these measures and it attenuated correlations.

Study One tested effectiveness of impression management scales to identify fakers. The faking identifier was dependent on opportunity to fake the measure. If an applicant was already high on the measure, he or she had no room to improve their score, thus they were not identified. However, they could be high on the measure because the

measure was transparent enough to fake. Impression management and the faking indicator are not able to estimate a true score for participant.

Future Research

In order to extend this research and address some of its limitations, personality measures that are more reliable could be used. Then accurate comparisons across applicant and incumbent conditions could be made. Measures that have a stronger relationship with job performance also should be used. Study Two looked at sales revenue and supervisor ratings, future research could look at different types of performance metrics like contextual performance, organizational citizenship behaviors. O'Connell, Kung, & Tristan (2011) found that SD measures are useful in correcting personality scores when performance is contextually based.

It would be interesting to conduct this research on personality measures that have items that are not as transparent to what they are measuring. Most of the participants in this study had scores at the ends of the distribution. In addition, personality measures that are less fakable are needed. Some such as, conditional reasoning measures (James, 1998) and forced-choice measures (Christiansen et al., 2005) have been shown to be less fakable, although there are concerns of their relationship to cognitive ability (Tristan et al., 2003; Christiansen et al., 2005).

Future research could also determine if re-test effects varied by subgroup, such as gender and age. In addition, personal characteristics that distinguish fakers from non-fakers, as well as personal circumstances (i.e., job desirability) that may lead someone to

fake could be measured. It is a generalization to assume that all applicants improve their score for the same reason and to the same degree.

The impression management scale used in these studies was based off of Paulhus' Balanced Inventory of Desirable Responding and the MMPI-Lie scale. These scales are used to detect socially desirable responding to a measure, not if you manage your impressions with others at work. It would be interesting to examine the convergent validity of our IM scale with an impression management scale that measures managing your impressions/behaviors with others socially. The relationship between the construct of impression management and performance could be re-examined by using a more construct valid measure.

Conclusion

This research was unique and extended the literature in several ways. Both studies had real-world samples, not university students taking the measures. Study One investigated faking with two types of indicators: an impression management scale and a faking indicator that detected score change from applicant and incumbent. Study One used a within-subjects design that had several advantages; in particular, it was able to isolate individual differences in faking across contexts without confounding effect of between-groups variance. Study One also addressed some of the limitations of Ellingson et al. (2007). There was no confound effect for time and order effects, and scales were used that are related to work performance. Study Two had objective and subjective job performance: weekly sales revenue and supervisor ratings. For Study Two, I looked at two ways that impression management scores are used in selection: removing applicants

from the selection pool based on high impression management scores or correcting personality scores to estimate a more honest score.

Impression management was able to identify fakers who had opportunity to improve their scores. However, not everyone fakes to the same degree and the same direction on all the personality measures. On average, those responding to the competitiveness scale as applicants lowered their score to appear more favorable and they were unable to be identified as fakers. Extraversion did not have a relationship to job performance in Study Two, perhaps because the extraversion scale was very transparent and easily faked.

Researchers have called for the faking and job performance relationship to be examined for sales jobs (Hogan, Hogan, & Roberts, 1996; Viswesvaran, Ones, & Hough, 2001). Study Two found that impression management does not have a relationship to job performance for salespeople. If in a real selection context impression management scales were used as a proxy for faking, removing applicants or correcting their scores based on impression management would improve job performance.

Table 1.
Study One: Score Differences between the Two Administrations.

	Administration #1		Administration #2		<i>t</i>	<i>d</i>
	M	SD	M	SD		
Group 1 (<i>N</i> = 3,136)						
Extraversion	18.30	3.16	18.28	3.19	.395	0.00
Competitiveness	4.71	2.06	4.67	2.08	1.03	0.02
Group 2 (<i>N</i> = 206)						
Extraversion	17.88	3.06	18.52	2.78	-2.57*	-0.22
Competitiveness	5.28	2.39	4.61	1.98	3.54**	0.31
Group 3 (<i>N</i> = 129)						
Extraversion	18.33	3.08	17.49	3.59	2.40*	0.25
Competitiveness	4.53	2.25	5.14	2.17	-2.47*	-0.28

Note. The groups comprise the archival dataset. Group 1: participants were applicants on the two administrations of the test. Group 2: participants were incumbents at time one and applicants at time two. Group 3: participants were applicants at time one and incumbents at time two.

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

Table 2.
Study One: Descriptive Statistics and Intercorrelations for Group One.

	Mean	SD	1	2	3	4	5	6	7	8
1. Sex	0.36	0.48	-							
2. Months between Tests	23.69	16.15	.01	-						
3. Extraversion Time 1	18.30	3.16	-.05**	-.01	-					
4. Extraversion Time 2	18.28	3.19	-.04*	.00	.45**	-				
5. Competitiveness Time 1	4.71	2.06	.12*	-.02	-.07**	-.05**	-			
6. Competitiveness Time 2	4.68	2.10	.13*	.03*	-.00	-.05**	.37*	-		
7. Impression Management Time 1	6.24	1.79	.05*	-.02	.23**	.12**	-.42**	-.22**	-	
8. Impression Management Time 2	6.27	1.80	.06**	-.01	.14**	.24**	-.24**	-.44**	.47**	-

Note. Group One participants were applicants on both administrations of the test. $N = 3,105$; Male = 0; Female = 1; * $p < .05$. ** $p < .01$.

Table 3.
Study One: Descriptive Statistics and Intercorrelations for Group Two

	Mean	SD	1	2	3	4	5	6	7	8
1. Sex	.28	.45	-							
2. Months between Tests	31.60	15.87	-.01	-						
3. Extraversion Time 1	17.88	3.06	.12	-.09	-					
4. Extraversion Time 2	18.52	2.78	.12	-.03	.26**	-				
5. Competitiveness Time 1	5.28	2.40	-.05	-.03	-.16*	-.02	-			
6. Competitiveness Time 2	4.61	1.99	-.06	-.05	-.01	-.04	.23**	-		
7. Impression Management Time 1	5.87	1.79	.11	-.01	.31**	.11	-.41**	-.19**	-	
8. Impression Management Time 2	6.21	1.77	.17*	-.01	.08	.34**	.08	-.45**	.29**	-

Note. Group Two participants were part of a validation study (incumbents) at time one and applicants at time two; $N = 205$; Male = 0; Female = 1; * $p < .05$. ** $p < .01$.

Table 4.
Study One: Descriptive Statistics and Intercorrelations for Group Three

	Mean	SD	1	2	3	4	5	6	7	8
1. Sex	.29	.45	-							
2. Months between Tests	19.55	14.99	-.03	-						
3. Extraversion Time 1	18.33	3.08	.09	-.04	-					
4. Extraversion Time 2	17.49	3.59	.03	.00	.29**	-				
5. Competitiveness Time 1	4.53	2.25	-.16	.04	.09	.16	-			
6. Competitiveness Time 2	5.14	2.17	-.04	.02	.08	.03	.19*	-		
7. Impression Management Time 1	6.36	1.62	.03	.02	.15	-.03	-.44**	-.12	-	
8. Impression Management Time 2	5.97	1.74	-.10	.13	.05	.20*	-.18*	-.48*	.23**	-

Note. Group Three participants were applicants on the first administration of the test and incumbents on the second. $N = 129$; Male = 0; Female = 1; * $p < .05$. ** $p < .01$.

Table 5.
Study One: Descriptive Statistics and Intercorrelations for Groups Two and Three Combined

	Mean	SD	1	2	3	4	5	6	7	8
1. Extraversion Incumbent	17.73	3.27	-							
2. Extraversion Applicant	18.45	2.89	.27**	-						
3. Competitiveness Incumbent	5.23	2.30	-.08	.02	-					
4. Competitiveness Applicant	4.58	2.08	.07	.02	.21**	-				
5. Impression Management Incumbent	5.91	1.77	.26**	.08	-.43**	-.19**	-			
6. Impression Management Applicant	6.27	1.71	.03	.26**	-.09	-.44**	.27**	-		
7. Extraversion Faker	0.25	0.43	-.68**	.14**	.03	-.08	-.21**	.12**	-	
8. Competitiveness Faker	0.21	0.41	.02	-.03	-.41**	.48**	.09	-.22**	-.02	-

Note. Group Two and Group Three ($N = 335$) were combined by their incumbent scores and applicant score for each of the personality measures.

Table 6.
Study One: Non-fakers and Fakers Means for Groups Two and Three.

	Non-Faker					Faker				
	Incumbent		Applicant		Cohen's D	Incumbent		Applicant		Cohen's D
	M	SD	M	SD		M	SD	M	SD	
Extraversion	19.00	2.13	18.21	3.14	.30	13.88	5.27	19.16	1.83	-1.49
Competitiveness	5.71	2.20	4.06	1.74	.84	3.36	1.70	6.55	2.17	-1.65

Note. There were 252 non-fakers on the extraversion measure and 83 fakers. There were 266 non-fakers on the competitiveness measure and 69 fakers. * $p < .05$. ** $p < .01$.

Table 7.
Study One: Classification of Extraversion Faker and Non-Faker based on the Discriminant Function.

Actual Group	Number of Cases	Predicted Group			
		Faker		Non-Faker	
		<i>n</i>	%	<i>n</i>	%
Faker	83	5	6%	78	94%
Non-Faker	250	4	1.6%	246	98.4%

Note. 74.4% of original grouped cases classified correctly.

Table 8.
Study One: Classification of Competitiveness Faker and Non-Faker based on the Discriminant Function.

Actual Group	Number of Cases	Predicted Group			
		Faker		Non-Faker	
		<i>n</i>	%	<i>n</i>	%
Faker	69	7	10.1%	62	89.9%
Non-Faker	264	1	0.4%	263	99.6%

Note. 81.1% of original grouped cases classified correctly.

Table 9.
Study Two: Descriptive Statistics and Intercorrelations

	<i>N</i>	Mean	S.D.	1	2	3	4	5	6	7	8	9
1. Sex	277	.17	0.37	-								
2. Age	266	26.08	5.52	-.01	-							
3. Tenure (weeks)	382	47.65	42.69	-.21**	-.06	-						
4. Weeks in Training	381	22.99	4.72	.07	.06	-.32**	-					
5. Extraversion	382	19.00	2.37	.09	.18**	-.07	.03	-				
6. Competitiveness	381	5.46	2.24	-.11	-.13**	.18**	-.14**	-.20**	-			
7. Impression Management	381	6.09	1.74	.03	.09	-.05	.11*	.29**	-.47*	-		
8. Weekly Sales Revenue	382	937.02	1421.77	-.20**	-.11	.81**	-.33*	-.04	.16**	-.07	-	
9. Supervisor Rating	382	2.19	0.48	-.15**	-.04	.86**	-.26**	-.06	.19**	-.04	.77**	-

Note. * $p < .05$, ** $p < .01$; Male = 0; Female = 1.

Table 10.
Study Two: Weekly Sales Revenue for Sample, Selected by Extraversion Scores.

Selection Ratio	Full Sample	Removed Top 6% of Impression Management Scores	Removed Top 24% of Impression Management Scores
.32	\$765.54	\$787.54	\$807.79
.55	\$799.71	\$806.47	\$830.47
.72	\$877.68	\$892.33	\$906.84

Table 11.
Study Two: Average Supervisor Ratings for Sample, Selected by Extraversion Scores.

Selection Ratio	Full Sample	Removed Top 6% of Impression Management Scores	Removed Top 24% of Impression Management Scores
.32	2.15	2.16	2.16
.55	2.17	2.18	2.18
.72	2.18	2.18	2.18

Table 12.
Study Two: Weekly Sales Revenue for Sample, Selected by Competitiveness Scores.

Selection Ratio	Full Sample	Removed Top 6% of Impression Management Scores	Removed Top 24% of Impression Management Scores
.11	\$1,687.17	\$1,687.17	\$1,687.17
.28	\$1,247.81	\$1,255.85	\$1,280.86
.64	\$1,031.64	\$1,042.93	\$1,045.94

Table 13.
Study Two: Average Supervisor Ratings for Sample, Selected by Competitiveness Scores.

Selection Ratio	Full Sample	Removed Top 6% of Impression Management Scores	Removed Top 24% of Impression Management Scores
.11	2.49	2.49	2.49
.28	2.32	2.32	2.31
.64	2.23	2.23	2.23

Table 14.
Study Two: Weekly Sales Revenue for Sample with Corrected Extraversion Scores.

Selection Ratio	Based on Extraversion Scores	Based on Extraversion Scores that were Corrected for Impression Management
.32	\$765.54	\$806.80
.55	\$799.71	\$829.80
.72	\$877.68	\$877.68

Note. There were 67 participants whose scores were corrected, thus the selection ratio for the top scores do not match between the uncorrected versus uncorrected. For the corrected scores, 20% of the participants had the top extraversion score of 21, 39% had scores of 21 or 20, and 72% had scores of 21, 20, or 19.

Table 15.
Study Two: Average Supervisor Ratings for Sample with Corrected Extraversion Scores.

Selection Ratio	Based on Extraversion Scores	Based on Extraversion Scores that were Corrected for Impression Management
.32	2.15	2.16
.55	2.17	2.18
.72	2.18	2.18

Note. There were 67 participants whose scores were corrected, thus the selection ratio for the top scores do not match between the uncorrected versus uncorrected. For the corrected scores, 20% of the participants had the top extraversion score of 21, 39% had scores of 21 or 20, and 72% had scores of 21, 20, or 19.

Table 16.
Study Two: Weekly Sales Revenue for Sample with Corrected Competitiveness Scores.

Selection Ratio	Based on Competitiveness Scores	Based on Competitiveness Scores that were Corrected for Impression Management
.11	\$1,687.17	\$1,687.17
.28	\$1,247.81	\$1,280.86
.64	\$1,031.64	\$1,031.64

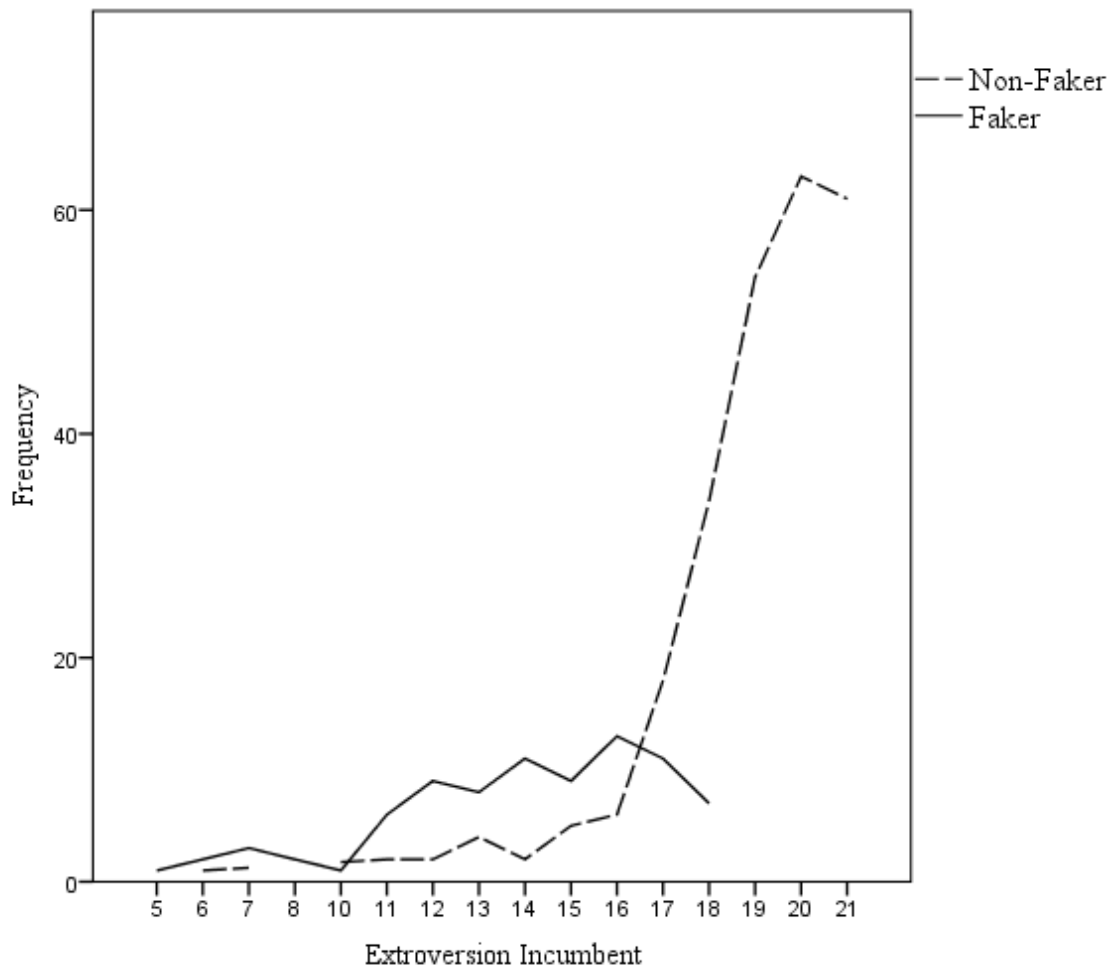
Note. There were 18 participants whose scores were corrected, thus the selection ratio does not match between the uncorrected and uncorrected at the middle selection ratio. For the corrected scores, 26% of the sample had competitiveness scores of 12, 11, 10, 9, 8, or 7 versus 28% for the uncorrected scores. The value for the selection ratio of .26 is in the second column.

Table 17.
Study Two: Average Supervisor Ratings for Sample with Corrected Competitiveness Scores.

Selection Ratio	Based on Competitiveness Scores	Based on Competitiveness Scores that were Corrected for Impression Management
.11	2.49	2.49
.28	2.32	2.31
.64	2.23	2.23

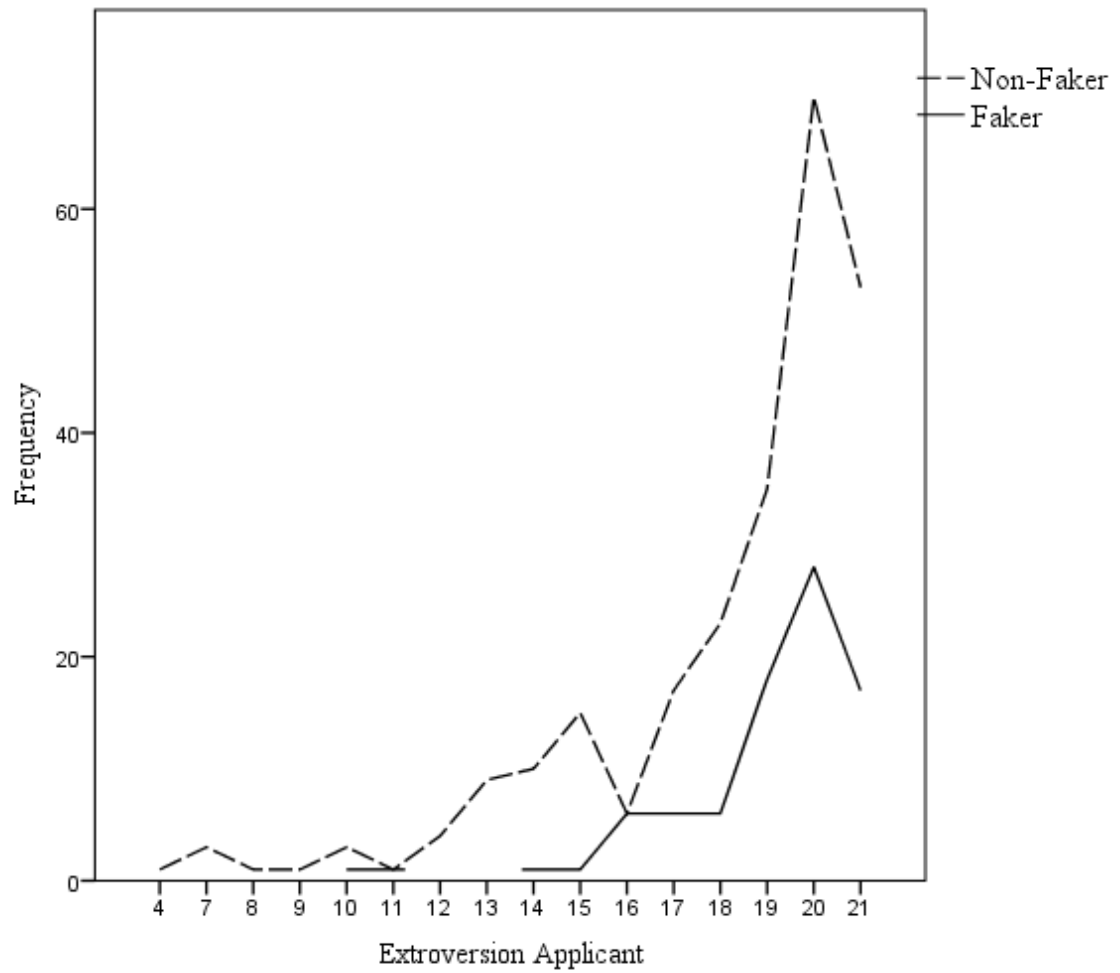
Note. There were 18 participants whose scores were corrected, thus the selection ratio does not match between the uncorrected and uncorrected at the middle selection ratio. For the corrected scores, 26% of the sample had competitiveness scores of 12, 11, 10, 9, 8, or 7 versus 28% for the uncorrected scores. The value for the selection ratio of .26 is in the second column.

Figure 1.
Extraversion Incumbent Scores.



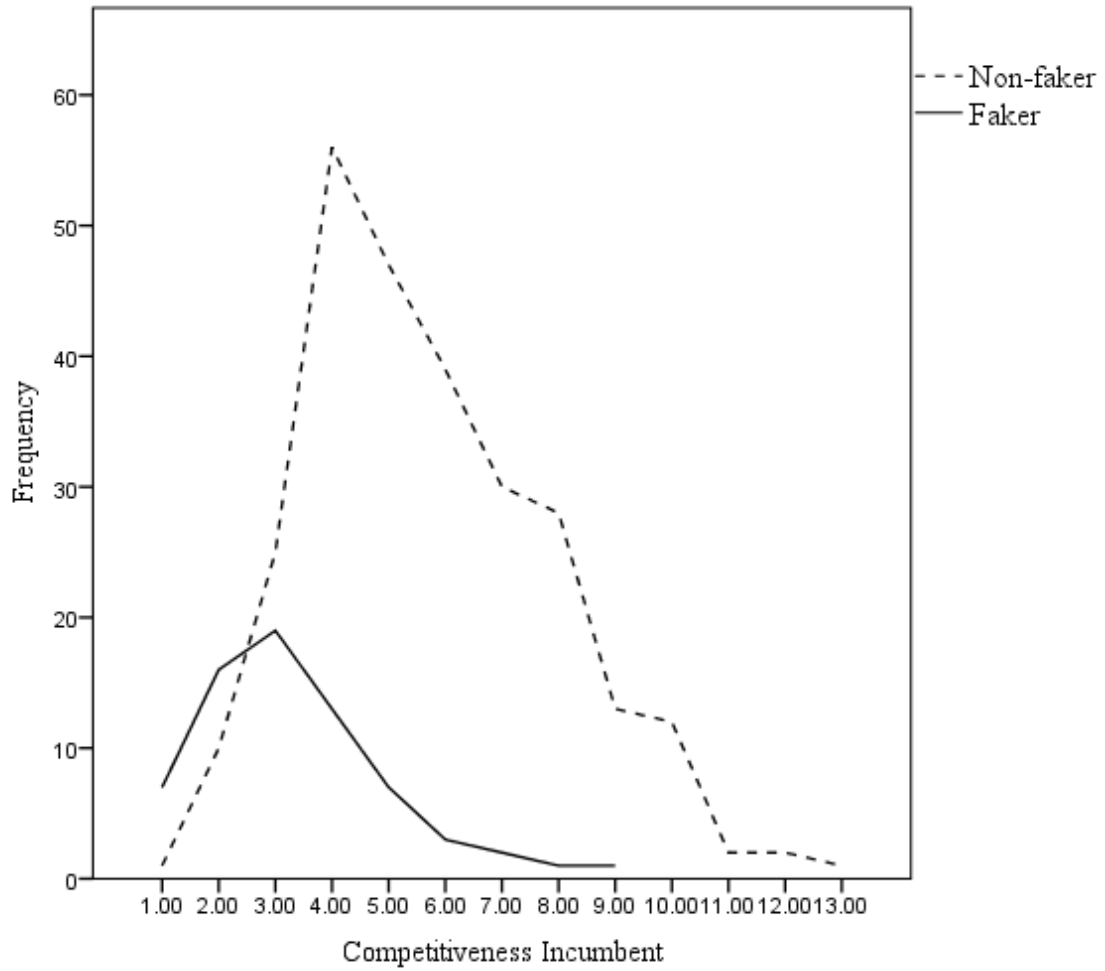
Note. Frequencies for Groups Two and Three ($N = 335$) are displayed.

Figure 2.
Extraversion Applicant Scores.



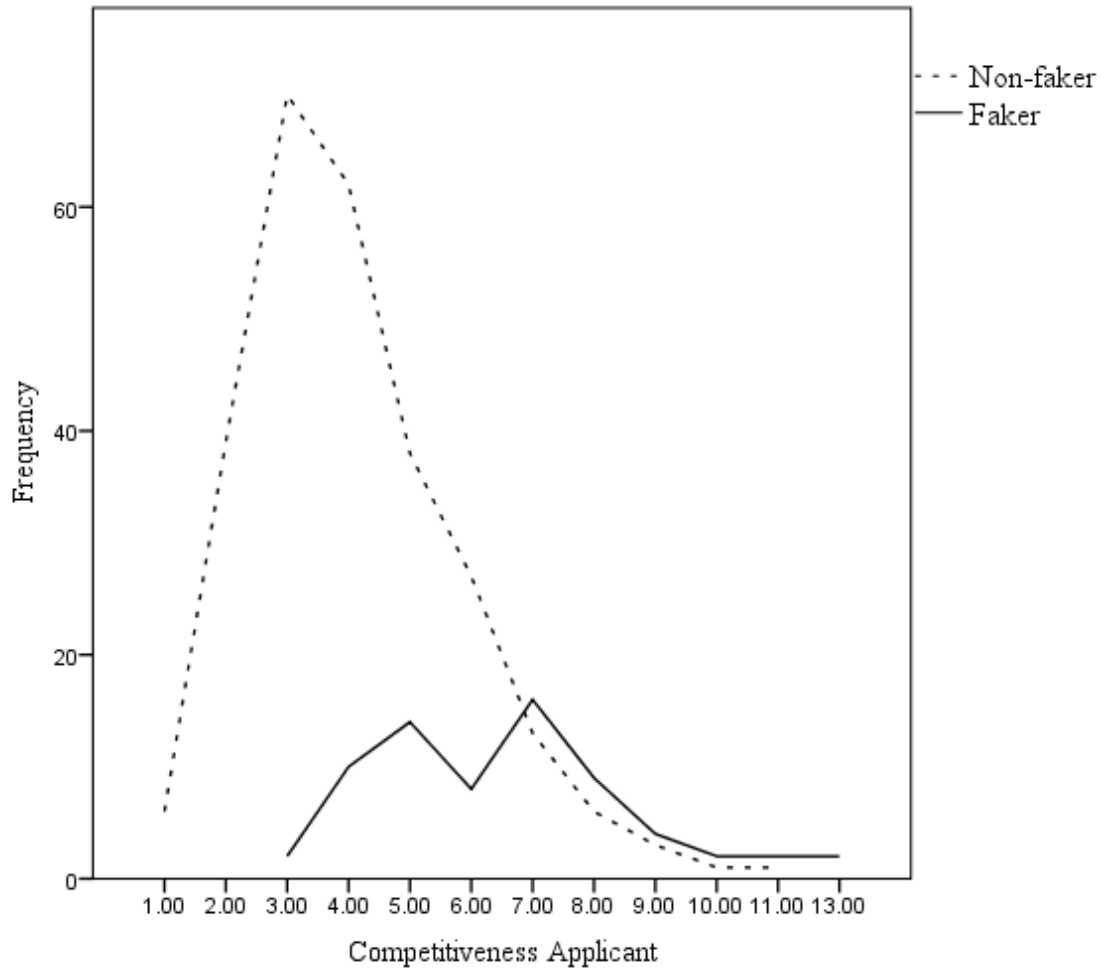
Note. Frequencies for Groups Two and Three ($N = 335$) are displayed.

Figure 3.
Competitiveness Incumbent Scores.



Note. Frequencies for Groups Two and Three ($N = 335$) are displayed.

Figure 4.
Competitiveness Applicant Scores.



Note. Frequencies for Groups Two and Three ($N = 335$) are displayed.

V. REFERENCES

- Alarcon, G. (2011). Social desirability and response distortion in conscientiousness scales. Poster at the 26th Annual Conference for the Society for Industrial and Organizational Psychology: Chicago, IL.
- Arthur Jr., W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2009). Unproctored internet-based tests of cognitive ability and personality: Magnitude of cheating and response distortion. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2(1), 39-45.
- Arthur Jr., W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18, 1-16.
- Avis, J.M., Kudisch, J.D., & Fortunato, V.J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology*, 17(1), 87-105.
- Azjen, I. (1991). Theory of planned behavior. *Organizational behavior and human Decision Processes*, 50, 179-211.

- Backstrom, M., Bjorklund, F., & Larsson, M.R. (2011). Social desirability in personality assessment: Outline of a model to explain individual differences. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp. 201 - 213). Oxford, England: Oxford University Press.
- Barrick, M.R. & Mount, M.K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*(3), 261-272.
- Barrick, M.R., & Mount, M.K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18* (4), 359-372.
- Barrick, M.R., Mount, M.K., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go from here? *Personality and Performance, 9* (1/2), 9-30.
- Barrick, M.R., Stewart, G.L., & Piotrowski, M. (2002). Personality and job performance: test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology, 87* (1), 43-51.
- Berry, C.M., Page, R.C., & Sackett, P.R. (2007). Effects of self-deceptive enhancement on personality- job performance relationships. *International Journal of Selection and Assessment, 15*(1), 94-109.
- Bing, M.N., Whanger, J.C., Davison, H.K., & VanHook, J.B. (2004). Incremental validity of the frame of reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150-157.

- Birkeland, S.A., Manson, T.M., Kisamore, J.L., Brannick, M.T., & Smith, M.A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Management, 14* (4), 317-335.
- Bott, J.P., O'Connell, M.S., Ramakrishnan, M., & Doverspike, D. (2007). Practical limitations in making decisions regarding the distribution of applicant personality test scores based on incumbent data. *Journal of Business Psychology, 22*, 123-134.
- Burns, G.N., & Christensen, N.D. (2006). Sensitive or Senseless: On the use of social desirability measures in selection and assessment. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 115-150). Greenwich: Information Age Publishing.
- Butcher, J.N., Williams, C.L., Graham, J.R., Archer, R.P., Tellegen, A., Ben-Porath, Y.S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory-Adolescent Version (MMPI-A): Manual for administration, scoring and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Christiansen, N.D., Goffin, R., Johnston, N., & Rothstein, M.G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*(4), 847-860.
- Christiansen, N.D., Burns, G.N., & Montgomery, G.E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18* (3), 267-307.
- Christiansen, N. D., Rozek, R. F., & Burns, G. (2010). Effects of social desirability scores on hiring judgments. *Journal of Personnel Psychology, 9*(1), 27-39.

- Converse, P.D., Oswald, F.L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2006). Forcing choices in personality measurement: Benefits and limitations. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 263-282). Greenwich: Information Age Publishing.
- Dilchert, S., & Ones, D.S. (2011). Application of preventative strategies. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp. 177-200). Oxford, England: Oxford University Press.
- Donovan, J.J., Dwight, S.A., & Hurtz, G.M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*, 81-106.
- Dwight, S.A., & Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1-23.
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Holt, Rinehart, and Winston, Inc.
- Ellingson, J.E. (2011). People fake only when they need to fake. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp.19-33). Oxford, England: Oxford University Press.
- Ellingson, J.E., Sackett, P.R., & Hough, L.M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84* (2), 155-166.
- Ellingson, J.E., Smith, D.B., & Sackett, P.R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*(1), 122-133.

- Ellingson, J.E., Sackett, P.R., & Connelly B.S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology, 92* (2), 386-395.
- English, A., Griffith, R.L., Graseck, M., & Steelam, L.A. (2005). Frame of reference, applicant faking and the predictive validity of non-cognitive measures: A matter of context. Manuscript submitted for publication.
- Fielder, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology, 27* (4), 307-316.
- Furman, A. (1986). Response bias, social desirability, and dissimulation. *Personality Individual Differences, 7* (3), 385-400.
- Furman, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: Research and Reviews, 9*(1), 46-55.
- Furnham, A., & Fudge, C. (2008). The five factor model of personality and sales performance. *Journal of Individual Differences, 29* (1), 11-16.
- Gammon, A.R., & Griffith, R.L. (2011). How do real applicants who are fakers compare to nonfakers? Poster at the 26th Annual Conference for the Society for Industrial and Organizational Psychology: Chicago, IL.
- Gawronski, B., LeBel, E.P., & Peters, K.R. (2007). What do implicit measures tell us? *Perspectives on Psychological Science, 2* (2), 181-193.
- Goffin, R.D., & Christensen, N.D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11* (1/2), 340-344.

- Griffith, R.L., & Converse, P.D. (2011). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp.34-52). Oxford, England: Oxford University Press.
- Griffith, R.L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An Examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341-355.
- Griffith, R., Malm, T., English, A., Yoshita, Y., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive biases, and individual differences. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior*. (pp. 115-150). Greenwich: Information Age Publishing.
- Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1(3), 308-311.
- Heggestad, E.D., Morrison, M., Reeve, C.L., & McCloy, R.A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91(1), 9-24.
- Henry, M., & Raju, N. (2006). The effects of traited and situational impression management on a personality test: An empirical analysis. *Psychology Science*, 48(3), 247-266.

- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*(1), 100-112.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92*(5), 1270-1285.
- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18*(4), 331-341.
- Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51* (5), 469-477.
- Holden, R.R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences, 44*, 311-321.
- Holden, R. R., & Book, A. S. (2009). Using hybrid rasch-latent class modeling to improve the detection of fakers on a personality inventory. *Personality and Individual Differences, 47*(3), 185-190.
- Holden, R. R., & Book, A. S. (2011). Faking does distort self-report personality assessment. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp.71-84). Oxford, England: Oxford University Press.
- Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49*(5), 446-450.

- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities, *Journal of Applied Psychology*, 75 (5), 581-595.
- Hough, L.M., Oswald, F.L., & Ployhart, R.E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment*, 9(1-2), 152-194.
- International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences (<http://ipip.ori.org/>). Internet Web Site.
- James, L.R. (1998) Measurement of personality with conditional reasoning. *Organizational Research Methods*, 1(2), 131-163.
- Johnson, J.A. & Hogan, R. (2006). A socioanalytic view of faking. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 209-232). Greenwich: Information Age Publishing.
- Komar, S., Brown, D.J., Komar, J.A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, 93(1), 140-154.
- Komar, S., Komar, J.A., Robie, C., & Tagger, S. (2010). Speeding personality measures to reduce faking: A self-regulatory model. *Journal of Personnel Psychology*, 9(3), 126-137.

- Kuncel, N.R., Borneman, M., & Kiger, T. (2011). Innovative item response process and Bayesian faking detection methods. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp.102-112). Oxford, England: Oxford University Press.
- Kuncel, N.R. & Borneman, M.J. (2007). Toward a new method of detecting deliberately faked personality: The use of idiosyncratic item responses. *Journal of Selection and Assessment*, 15(2).
- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2010). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 6 (1), 202-210.
- Landy, F. J. (2008). Stereotypes, Bias and Personnel Decisions: Strange and Stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1 (4), 379-392.
- LeBreton, J.M., Barksdale, C.D., Robin, J., & James, L.R. (2007). Measurement issues associated with conditional reasoning: Indirect measurement and test faking. *Journal of Applied Psychology*, 92 (1), 1-16.
- Levashina, J., Morgeson, F. P., & Campion, M. A. (2009). They don't do it often, but they do it well: Exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Assessment*, 17(3), 271-281.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14(2), 131-141.

- Maccann, C., Ziegler, M., & Roberts, R.D. (2011). Faking in personality assessment: Reflections and recommendations. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp. 309-329). Oxford, England: Oxford University Press.
- Martin, B.A., Bowen, C.C., & Hunt, S.T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, 32, 247-256.
- McFarland, L.A., Ryan, A.M., & Ellis, A. (2002). Item placement on personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78(2), 348-369.
- McFarland, L.A., & Ryan A.M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology*, 36(4), 979-1016.
- Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review of research designs and analytic strategies. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 85-113). Greenwich: Information Age Publishing.
- Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: The role of over-claiming, self-esteem, and emotional intelligence. *Psychology Science*, 48(3), 336-356.
- Miller, C. E., Tristan, E., Leasher, M. K., & Thompson, C. N. (2005). Validation of the Chally Assessment System with a sample of sales representatives. *Applied H. R. M. Research*, 10, 39-44.

- Morgeson, F.P., Campion, M.A., Diboye, R.L., Hollenbeck, J.R., Murphy, K., & Schmidt, N. (2007). Reconsidering the use of personality tests in selection contexts. *Personnel Psychology, 60*, 683-729.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.
- O'Connell, M.S., Kung, M., & Tristan, E. (2011). Beyond impression management: Evaluating three measures of response distortion and their relationship to job performance. *International Journal of Selection and Assessment, 19* (4), pagination not specified.
- Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for Personnel Selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.
- Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.
- Paulhus, D.L. (1998). *Manual for the Paulhus Deception Scales: BIDR Version 7*. Toronto: Multi-Health Systems.
- Paulhus, D.L., Harms, P.D., Bruce, M.N., & Lysy, D.C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*(4), 890-904.
- Pauls, C., & Crost, N.W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37*, 1136-1151.

- Peterson, M.H., & Griffith, R.L. (2006). Faking and job performance: A multifaceted issue. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 223-261). Greenwich: Information Age Publishing.
- Peterson, M. H., Griffith, R. L., & Converse, P. D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single- and multiple-predictor selection. *Journal of Business and Psychology*, 24(4), 373-386.
- Peterson, M.H., Griffith, R.L., Converse, P.D., & Gammon, A.R. (2011). Using within-subjects designs to detect applicant faking. Paper presented in Converse, P.D. (2011, April) *Detecting deception: Techniques for assessing applicant faking on personality measures*. Symposium conducted at the 26th Annual Conference for the Society for Industrial and Organizational Psychology: Chicago, IL.
- Reeder, M.C., & Ryan, A.M. (2011). Methods for correcting for faking. In M. Ziegler, C. Maccann, R.D. Roberts (Eds.) *New perspectives on faking in personality assessment*. (pp. 131-150). Oxford, England: Oxford University Press.
- Robie, C., Brown, D.J., & Beaty, J.C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, 21(4), 489-509.
- Robinson, M.D., & Neighbors, C. (2006). "Catching the mind in action: Implicit methods in personality research and assessment." In M. Eid and E. Diener (Eds.) *Handbook of multi-method measurement in psychology* (pp. 115-125). Washington, D.C.: American Psychological Association.

- Robinson, M.D., & Neighbors, C. (2006). "Catching the mind in action: Implicit methods in personality research and assessment." In M. Eid and E. Diener (Eds.) *Handbook of multi-method measurement in psychology* (pp. 115-125). Washington, D.C.: American Psychological Association.
- Roth, P.L., Bevier, C.A., Bobko, P., Switzer III, F.S., & Tyler P. (2006). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54 (2), 297-330.
- Ruch, F.L., & W.W. Ruch. (1967). The K factor as a (validity) suppressor variable in predicting success in selling. *Journal of Applied Psychology*, 51(3), 201-204.
- Ryan, A.M., & Boyce, A.S. (2006). What do we know and where do we go? Practical directions for faking research. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 357-371). Greenwich: Information Age Publishing.
- Rynes, S.L. (1993). "Who's selecting whom? Effects of selection practices on applicant samples and behavior", In N. Schmitt and W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 240-274). San Francisco: Jossey Bass.
- Schleicher, D.J., Van Iddekinge, C.H., Morgeson, F.P., & Campion, M.A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology*, 95(4), 603-617.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implication of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.

- Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on measures in selection settings. *Journal of Applied Psychology, 91*(3), 613-621.
- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*, 979-988.
- Sisco, H., & Reilly, R. R. (2010). Effect of item level social desirability on factor structure stability. *International Journal of Selection and Assessment, 18*(3), 305-309.
- Smith, D.B., & Ellingson, J.E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*(2), 211-219.
- Snell, A.F., & Fluckinger, C.D. (2006). Understanding responses to personality selection measures: A conditional model of the applicant reasoning process. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 179-208). Greenwich: Information Age Publishing.
- Snell, A.F., Sydell, E.J., & Lueke, S.B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resources Management Review, 9* (2), 219-242.
- Stark, S., Chernyshenko, O.S., Chan, K., Lee, W.C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86*, 943-953.
- Steffens, M.C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*(3), 165-179.

- Steffens, M.C., & Konig, S.S. (2006). Predicting spontaneous Big Five behavior with Implicit Association Tests. *European Journal of Psychological Assessment, 22*(1), 13-20.
- Stewart, G. L., Darnold, T. C., Zimmerman, R. D., Parks, L., & Dustin, S. L. (2010). Exploring how response distortion of personality measures affects individuals. *Personality and Individual Differences, 49*(6), 622-628.
- Tett, R.P., Anderson, M.G., Ho, C., Yang, T.S., Huang, L., & Hanvongse, A. (2006). Seven nested questions about faking on personality tests: An overview and interactionist model of item-level response distortion. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 43-83). Greenwich: Information Age Publishing.
- Tristan, E., Miller, C.E., & Leasher, M.K. (2003). Conditional reasoning tests of achievement orientation: Fakability and adverse impact. Poster presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, Florida, April, 2003.
- Tristan, E. (2009). Measuring applicant faking with job desirability: Prevalence, selection, and measurement issues in an applied sample. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 70*(4-), 2605.
- Vasilopoulos, N.L., & Cucina, J.M. (2006). Faking on non-cognitive measures: The interaction of cognitive ability and test characteristics. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior* (pp. 305-331) Greenwich: Information Age Publishing.

- Vasilopoulos, N.L., Reilly, R.R., & Leaman, J.A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology, 85*, 50-64.
- Van Iddekinge, C.H., Morgeson, F.P., Schleicher, D.J., & Campion, M.A. (2011). Can I retake it? Exploring subgroup differences and criterion-related validity in promotion retesting. *Journal of Applied Psychology*, No Pagination Specified. doi: 10.1037/a0023562
- Viswesvaran, C., & Ones, D.S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210.
- Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment, 9*(4), 277-289.
- Waters, L.K. (1965). A note on the “fakability” of forced-choice scales. *Personnel Psychology, 18*, 187-191.
- Zickar, M.J., & Gibby, R.E. (2006). A history of faking and social desirable responding on personality tests. In R.L. Griffith & M.H. Peterson (Eds.) *A closer examination of applicant faking behavior*. (pp. 21-42) Greenwich: Information Age Publishing.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item level analysis. *Journal of Applied Psychology, 84*, 551-563.
- Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement, 69*(4), 548-565.

Ziegler, M. (2011). Applicant faking: A look into the black box. *The Industrial-Organizational Psychologist*, 49 (1), 29-36.