2010

Measuring Conscientiousness with Explicit and Implicit Measures

Jenna Noelle Filipkowski
Wright State University

MEASURING CONSCIENTIOUSNESS WITH EXPLICIT AND IMPLICIT
MEASURES

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

By

JENNA NOELLE FILIPKOWSKI

B.S., Ursinus College, 2007

2010
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

<u>November 20, 2009</u>

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY
SUPERVISION BY <u>Jenna Noelle Filipkowski</u> ENTITLED <u>Measuring Conscientiousness
with Explicit and Implicit Measures</u> BE ACCEPTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF <u>Master of Science</u>.

_____
Corey E. Miller, Ph.D.
Thesis Director

_____
Director, HF-IO Psychology
Ph.D. Program

Committee on
Final Examination

_____
Debra Steele-Johnson, Ph.D.

_____
Gary Burns, Ph.D.

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

ABSTRACT

Filipkowski, Jenna Noelle. M.S., Department of Psychology,
Industrial and Organizational Psychology Program, Wright State University, 2010.
Measuring Conscientiousness with Explicit and Implicit Measures

The fakability of three measures of conscientiousness was examined: the International

personality Item Pool (IPIP), the Conditional Reasoning Test, and Implicit Association

Tests (IAT). Data from a student sample ($N = 442$) found the Conditional Reasoning and

IATs were the least susceptible to faking, but they did not have a meaningful relationship

with the IPIP. The Conditional Reasoning Test was the best non-cognitive predictor of

College GPA ($r = .23, p < .01$), perhaps due to the fact it may tap cognitive ability as

evidenced by its significant correlation with the ACT ($r = .41, p < .01$). The explicit (self-

report) and implicit (IAT) measures of conscientiousness had a low positive correlation

with one another. Some researchers may use this finding as evidence that they are

measuring different aspects of the construct.

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

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 spend their time and money wisely on testing and selection. On the other hand, most applicants want to be hired. 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, which is a concern to employers and Industrial/Organizational psychologists. If applicant faking on personality measures occurs and it negatively affects selection, researchers need to determine what the best way is to deter or prevent faking behavior.

The literature on applicant faking on personality measures is extensive. This introduction will review the literature and address seven topics: 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, controlling applicant faking, and using implicit measures to control for 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 (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 ($\rho = .27$). Similar to conscientiousness, emotional stability was a valid predictor of work performance across jobs ($\rho = .13$), and emotional stability was a valid predictor of teamwork ($\rho = .22$). Extraversion correlated significantly 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 group; 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 certain jobs. In a sample of customer-service employers, conscientiousness predicted job performance better than cognitive ability (Avis, Kudisch, & Fortunato, 2002). Another meta-analysis that used 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 .34 for Agreeableness to .43 for Emotional Stability. Researchers also have suggested a personality composite is a better predictor than individual personality traits in isolation (e.g., 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. Cognitive ability measures

are the best predictors of job performance; however, they exhibit adverse impact, i.e., score differences for racial groups (Hunter & Hunter, 1984). 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 (e.g., 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 an operational definition of faking (Zickar, Gibby, & Robie, 2004). The terms used for faking are response distortion, impression management, social desirability, self-enhancement, and claiming unlikely virtues (Griffith, Chemielowski, & 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. 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, Morrson, Reeve, & McCloy, 2006).

Researchers have not agreed on a universal definition of faking, nor have they agreed on a theoretical model of faking behavior. There are four theoretical models of applicant faking. Snell, Sydell, and Lueke (1999) developed an interactional model of faking, in which they posited that 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 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).

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. Their interactive model is based on classic true score theory. 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 this 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). Justification mechanisms are a part of the reasoning process used by responders when selecting certain items based on their latent motives. There are five justification mechanisms for applicant responding: verification (consequences of faking), utility (incentives to fake), morality (dispositions of morals and values), ability (knowledge, skills, and abilities a for target job), and framing (framing a response to a situation). 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 influence an applicant's intention to fake. Situational factors such as an incentive for doing well on the test and warnings not to fake moderate the intention 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).

The theories of applicant faking have a common tenet as they all describe faking behavior as a complex interaction between personality, situations, and ability. Researchers have 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. 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.

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 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 job 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 mangers from several organizations (Martin, Bowen, & Hunt, 2002).

These studies illustrate 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 be 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).

Researchers use many research designs to determine the prevalence of faking behavior (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 present 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. 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 is a high score on a social desirability measure in a within-subjects design. Another within-subjects design is to examine personality score differences of an individual as an applicant then as an employee of an organization (see Mesmer-Magnus & Viswesvaran, 2006).

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 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).

Other studies have used within-subjects designs to determine if faking behavior occurs. 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. Griffith, Chemielowski, and Yoshita (2007) examined the prevalence of applicant faking in 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 $M = 191.79$, $SD = 27.23$]. They also found that the rank ordering of applicants changed when response distortion occurred (Griffith et al., 2007).

Not all researchers believe that faking occurs. Hough, Eaton, Dunnette, Kamp, and McCloy (1990) reported that faking might not occur in real-world 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).

Researchers conducting studies on real-world applicants determined 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. 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.

Ellingson, Sackett, and Connelly (2007) also used a within-subjects design and found response distortion is not the norm. Participants took personality measures as applicants then later as incumbents for developmental purposes. The researchers found that applicants engaged in a limited amount of response distortion. The researchers noted some limitations in their study. Researchers used the California Personality Inventory as the personality measure, and the California Personality Inventory is made of subtle items that may be less susceptible to intentional response distortion. 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 have found evidence of faking in laboratory studies (e.g., Viswesvaran & Ones, 1999; Mueller-Hanson, Heggestad, & Thornton, 2003) and in field studies (e.g., Birkeland et al., 2006; Griffith, Chemielowski, & 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, define faking as scoring high in social desirability. The evidence seems to suggest that faking on personality measures is possible and occurs within selection. The studies that have contradictory findings have several different methodological problems.

The different samples in these types of investigations may explain why the research on the prevalence of faking is mixed (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). Also researchers have attributed the discrepancies in the faking literature to the ways faking has been operationalized (McFarland & Ryan, 2006). Most researchers have used the terms faking and social desirability interchangeably; however, they are two different concepts. Applicant faking is having false responses on a personality measure to create a specific impression, good or bad. Social desirability refers to one specific type of faking, presenting one's self in a favorable light (Furman, 1986). Individual differences and situational demands influence faking (McFarland & Ryan, 2006) whereas social desirability is more trait-like (Furman, 1986; Morgeson, Campion, Diboye, Hollenbeck, Murphy, & Schmidt, 2007; Ones, Viswesvaran, and Reiss, 1996).

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

11

important concern to faking researchers. The impact of faking on the selection process and hiring decisions may be more useful for understanding this process and for conducting future research.

The Impact of Applicant Faking on Selection

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-subject 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. Honest group applicants were under-selected in all the selection ratios compared to faking group applicants. 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).

Researchers also have studied the effects of faking on criterion-related validity. In a recent study, researchers conducted a Monte Carlo investigation of the effects of faking on criterion-related validity (Komar, Brown, Komar, & Robie, 2008). 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).

Another study found that using incumbent scores for cutoffs in selection costs the organization extra money in the selection process (Bott, O'Connel, 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. They 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).

Applicant faking on personality measures has 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). Research has shown that fakers can be low performers, high performers, or equal performers (Vasilopoulos & Cucina, 2006). More research is needed addressing the job performance of fakers, in particular how faking affects contextual performance and counterproductive work behaviors.

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 review, see Zikar & Gibby, 2006). Researchers have used various methods and measures to detect applicant faking and to determine the criterion-related validity of the personality measure. Research has used a fake-good condition versus honest condition and applicants versus incumbents, and condition type could be a within-subjects design or a between-subjects design. Other conditions have included face-to-face interaction versus computer and anonymous versus non-anonymous administrations (Mesmer-Magnus & Viswesvaran, 2006). The most popular and oldest method for detecting fakers is the use of social desirability scales. Other methods for faking detection are verbal protocol analysis, reporting response latencies, idiosyncratic response patterns, bogus items, the randomized response technique, and item response theory. The methods of faking detection will be discussed below.

Social desirability scales. The first conceptualization of social desirability was as a response style not a substantive trait (Burns & Christensen, 2006). In fact, the social desirability measures were created before the construct was defined (Burns & Christensen, 2006). Edwards (1957) created the first 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 -.50 with Edward's social desirability scale). In 1960, researchers created the Marlowe-Crowne social desirability scale to address the limitations in Edward's scale (Zikar & Gibby, 2006).

Paulhus (1984) found that social desirability scales do not correlate well with each other. Marlowe-Crowne's and Edwards' social desirability scales had a low correlation of .24. Paulhus (1984) examined the scale item loadings 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).

Paulhus (1984) found that impression management scales can be faked; however, self-deceptive social desirability are not faked as much. However, recent research found both the impression management scale and the self-deceptive enhancement 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). This study examined mean differences in a within-subjects design of instructions to fake good (IM scores $M = 5.37$, $SD = 1.04$; self-deception score $M = 5.20$, $SD = 0.74$) or respond honestly (IM scores $M = 3.12$, $SD = 0.87$; self-deception 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 be concerned also 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. Nonetheless, researchers still correct personality scores based on social desirability scales (Goffin & Christensen,

2003). One way to correct for faking based on social desirability scale scores is to disregard the personality tests of applicants with high social desirability scores. Another option is to lower the 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 score with a special equation (Goffin & Christensen, 2003).

Some researchers believe corrections for faking are frivolous. Making corrections does not increase the validity of the test or increase mean performance (Schmitt & Oswald, 2006). Using social desirability scores to correct personality scores fails to produce a score that approximates the 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). Other research has 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 (e.g., Mueller-Hanson et al., 2003). Holden (2007) found that a moderated regression analysis using social desirability scales underestimated the effect sizes for induced faking on validity.

The use of social desirability scales to detect and correct for faking is controversial, yet such corrections 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 or lie scales within them. Such tests include the California Personality Inventory, Minnesota Multiphasic Personality Inventory, and 16 Personality Factor. Goffin and Christensen (2003) mailed a survey to 67 I/O psychologists, asking them if they used a response validity scale when they administered

personality measures for selection purposes. Thirty-six responded of whom 56%

indicated that they used a personality measure that included a response validity scale.

However, a high score on a social desirability scale does not always mean that

someone is likely to fake. Social desirability is a dispositional variable, and the faking

models mentioned above show that faking behavior is a function of dispositions and

situations. In their review of the social desirability research, Burns and Christensen

(2006) highlighted some of the unanswered questions about social desirability scales and

called for more research to be conducted. Some of the "unanswered" areas involve the

relationship between social desirability scores and actual 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). In addition, future research should determine if the social desirability

scales are all measuring the same construct and if they are highly correlated with each

other. Despite the large amount of research on social desirability measures, the field is far

from consensus. Rather than continue this particular line of research, I have decided to

explore a new test method; therefore, social desirability measures are beyond the scope of

the proposed research.

Other faking detection methods. The most direct method for detecting faking is

using a verbal protocol analysis. The 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 verbalized their thoughts, recorded by a tape recorder. 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: 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 fakers. However, the results have been mixed. Some studies have found fakers have slower response times whereas other research has found fakers to respond faster (for a review, see Vasilopoulos, Reilly, & 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 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 captured the rating and the response latency. The 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).

Another technique for detecting faking is using idiosyncratic response patterns (Kuncel & Borneman, 2007). In their study, 215 undergraduates completed 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 in an 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 the 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).

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, events, etc. 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).

Finally, 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 an IRT analysis to see the 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 between fakers and honest respondents differed across the personality scales and conditions. Researchers have pointed out some problems with using IRT to detect faking (Mesmer-Magnus & Viswesvaran, 2006). 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).

In sum, social desirability scales were the first method of detecting faking, and they still complement personality measures today. However, applicant faking is more complicated than social desirability, and these scales are not good indicators of faking behavior. The other methods for detecting faking are more useful in laboratory studies than actual applicant selection. Instead of using controversial methods to detect faking, researchers should develop measure 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. 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 predictors of outcomes 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.

 Donovan, Dwight, and Hurtz (2003) used the randomized-response technique to assess the prevalence of faking. Their study utilized student participants. Participants were less likely to engage in faking behaviors that were high in severity and verifiability (Donovan et al., 2003).

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 (e.g., McFarland, Ryan, & Ellis, 2002). One study found that the grouped construct format was more fakable than randomized format for personality scales measuring neuroticism and conscientiousness (McFarland et al., 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; 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 ("I will find out that you are faking, and you get in trouble") influences the effectiveness. Meta-analytic findings showed that identification-warnings had a $d = .01$, consequence-warning had a $d = .30$, and both types of warnings together had a $d = .25$ (Dwight & Donovan, 2003).

McFarland and Ryan (2006) found warnings had a direct effect on intention to fake and actual faking behavior. Warnings may not be effective for all respondents; warnings may cue risk-takers to attempt faking (Tett et al., 2006). Warnings may not be useful in the long run as test-takers catch on to the method (Zikar & 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 (Zikar & Gibby, 2006). Christiansen, Burns, and Montgomery (2005) conducted three studies on forced-choice items and their susceptibility to response distortion. In study 1, undergraduates ($N = 350$) completed personality measures including a forced-choice format and then completed more measures in either an honest or applicant condition. Scores were higher in the applicant condition ($d = .43$, $F(1, 348) = 21.30$, $p < .01$) than in the honest condition. However, scores were even higher in the applicant condition in the traditional personality format ($d = .71$, $F(1, 348) = 68.43$, $p < .01$). Social desirability scores from applicant conditions explained 35% of the variance on the traditional personality format scores, whereas forced-choice scores explained 1% of the variance in

social desirability scores. Study 2 found that forced-choice method was a better predictor of supervisors' ratings of performance than the other method. Study 3 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).

Some researchers have noted the limitations of forced-choice measures (for a review, see Converse, Oswald, Imus, Hedricks, Roy, & Butera, 2006). Some forced-choice items are ipsative, measuring intraindividual differences and not interindividual. Interindividual 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. Heggestad 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 (Zikar & Gibby, 2006).

The final and most recent method for controlling for faking on personality measures is the use of a conditional reasoning measure. 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 assumed that 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 1, comprised of undergraduate students, found those in the experimental group (those who were informed of 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 2 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 3 examined the scores of job applicants, job incumbents, and undergraduates, and 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) (LeBreton et al., 2007).

As LeBreton et al. (2007) demonstrated in their study, the conditional reasoning measure was insusceptible to faking. Respondents answered the test to find the correct and most logical answer and did not think it was measuring their personality. LeBreton et al. (1997) used a conditional reasoning test for achievement motivation, which is a facet of the construct of conscientiousness, and found that the conditional reasoning achievement instrument is unfakable.

There have not been many conditional reasoning tests created. James, McIntyre, Glisson, Bowler, and Mitchell (2004) created the first conditional reasoning measure. The conditional reasoning aggression scale has reliability of .76 and an average validity of .44, and it has been effective in predicting counterproductive work behaviors. Other conditional reasoning measures of team orientation, anti-social personality, and social bias

have been 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 measures is onerous (Robie et al., 2007).

Besides being very difficult to develop, conditional reasoning tests have demonstrated another drawback: they may show adverse impact as a selection measure towards minorities. Tristan, Miller, and Leasher (2003) found in a within-subjects design that the conditional reasoning measure 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).

In sum, each of the methods and measures for controlling has faking has advantages and disadvantages. Vasilopoulous and Cucina (2006) believed that methods for controlling faking may have introduced a cognitive aspect into the test that should not be there. This is a problem because one of the reasons personality measures are used in selection is because they do not exhibit the adverse impact observed for 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.

Implicit Measures of Personality

Implicit personality measures such as the Implicit Association Test can accomplish both of the above goals. Implicit Association Test research has begun to be done in the subfield of Industrial/Organizational Psychology, admittedly with controversy (Landy, 2008). There are many measures of implicit cognitions such as the Thematic Apperception Test, attributional style questionnaires, conditional reasoning measures, and implicit association tests (IAT) (for a review, see Bing, LeBreton, Davison, Migetz, & James, 2007). Implicit cognitions are unconscious and automatic thoughts that an individual has about his or her behavior, needs, values, affect, cognition, self-concept, and attitudes (see Bing et al., 2007). Researchers have differences labeling either measurement methods or constructs as implicit versus explicit (Banse & Greenwald, 2007). In this paper, implicit personality refers to a personality construct measured with an implicit measure, such as the IAT. An implicit measure is an assessment that reflects the to-be-measured construct by processes that are uncontrolled, unintentional, goal independent, purely stimulus driven, autonomous, unconscious, efficient, or fast (De Houwer & Moors, 2007).

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 measures of an individual's subjective trait level. Implicit measures are based on performance, and they capture the mind in action. Mental events take time and are measured chronometrically. How fast someone is able to categorize a concept illustrates how accessible that concept is in his or her mind. An accessible concept is one that is habitually used, and the respondent is able to make a quick

association and response (Robinson & Neighbors, 2006). Implicit measures use activation processes whereas explicit measures use validation processes (Gawronski, LeBel, & Peters, 2007).

The implicit association test. How quickly a respondent is able to make an association and a response is the rationale for the implicit association test (e.g., Steffens & Konig, 2006). Researchers first used IATs to assess implicit attitudes and prejudices; however, there are now IATs for self-esteem and personality self-concept (for a review, see Asendorph, 2007). 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 between himself or herself and being conscientiousness.

A standard IAT procedure contains seven blocks of trials (Schnabel, Asendorph, & Greenwald, 2007). Below is an example of an IAT measuring conscientiousness. Block 1 is a target-discrimination task with 40 trials, and the stimulus words are the terms for "self" or "other". The left key is for self and the right key is for other. Block 2 is an attribute-discrimination task with 40 trials, and the stimulus words are the conscientiousness terms and the non-conscientiousness terms (or negligence). The left key is for conscientiousness and the right key is for negligence. Block 3 is a combined-discrimination task with 40 trials. The left key is for conscientiousness + self, and the right key is for negligence + other. In the combined tasks, the stimulus target words are "self" words, "other" words, conscientiousness, and negligence terms. Block 4 is also a combined-discrimination task with 80 trials. The left key is for conscientiousness + self, and the right key is for negligence + other. Block 5 is an attribute-discrimination reversed task with 40 trials. The left key is for negligence, and the right key is for

conscientiousness. Block 6 is a combined-discrimination task with 40 trials. The left key is for negligence + self, and the right key is for conscientiousness + other. Block 7 is a combined-discrimination task with 80 trials. The left key is for negligence + self, and the right key is conscientiousness + other (Schnabel, Asendorph, & Greenwald, 2007; Grumm & Collani, 2007). Someone who is high in conscientiousness, will have faster response times in blocks 3 and 4, and will have slower response times in blocks 6 and 7. The reverse is true for someone who is not conscientious.

IAT psychometrics and validity. The IAT procedure displays adequate reliability. Cronbach's alpha of IATs show values in the range of .80 (Egloff & Schmukle, 2002). Across twenty studies with the IAT, test-retest reliability had a mean of .50 (Lane, Banaji, Nosek, & Greenwald, 2007). However, internal consistency and test-retest reliability of IATs are lower than explicit self-report measures.

A review of 122 research reports found average $r = .274$ for the prediction of behavioral, judgment, and physiological outcomes by IAT measures. The review found that for socially sensitive topics, predictive validity of self-report measures was low, and incremental validity of IAT measures was relatively high (Greenwald, Poelman, Uhlmann, & Banaji, 2009).

One particular study tested the validity of the Big Five personality traits assessed with an IAT (Schmukle, Back, & Egloff , 2008). They tested the factorial structure of an IAT approach using confirmatory factor analyses. They created subtests of items for each of the five factors. For example, neuroticism had five IATs, and the stimuli were anxious-calm, nervous-relaxed, fearful-restful, uncertain-at ease, and afraid-balanced. Each Big Five factor had five IATs, and each IAT loaded on the correct Big Five factor. The factor

loadings had a mean of .57. The correlations between the IATs and self-report measures was low or non-significant. The correlation between the conscientiousness IATs and self-report was low but significant ($r = .26$, $p < .05$). Researchers also examined explicit and implicit personality measures correlations with reported behavior. Participants with higher implicit conscientiousness scores, and controlling for the effects of self-reported conscientiousness and sex, reported cleaning their windows more frequently.

Most IAT studies examine the correspondence between explicit and implicit measures. A meta-analysis of IAT and self-report correspondence found an average correlation of .19 (Hoffman, Gawronski, Gschwendner, Le, & Schmitt, 2005). Several moderators influence the strength of explicit and implicit correspondence (for a review, see Hoffman, Gschwender, Nosek, & Schmitt, 2007). Moderators of explicit and implicit correspondence are the conceptual similarity between measures, spontaneity of the verbal responses, self-presentation tendencies, and evaluative strength.

Gschwedner, Hoffman, and Schmitt (2008) found that the level of content similarity and specificity moderated the relationship between explicit and implicit measures of anxiety. They also found that explicit and implicit measures of anxiety both accounted for unique variance in predicting anxious behavior while giving a speech. There was a relationship between the two types of measures whether they were assessing the same content, which was spider or speech anxiety. The results of a confirmatory factor analyses showed that explicit and implicit measures of anxiety factors correlated significantly on the same specification level, i.e., general-factor anxiety or specific-factor anxiety (Gschwedner, Hoffman, & Schmitt, 2008).

Fakability of the implicit association test. After reviewing the current research on the deliberate attempt to control responses on implicit measures, researchers concluded that these measures are not immune to this behavior (Gawronski, LeBel, & Peters, 2007). IAT scores can be altered deliberately, especially when respondents are given strategies to improve their scores. 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. The IAT scores fell in a negative direction in the second administration of the test. Successful faking occurred when participants were told to think about what strategies would increase their score and when participants were told explicitly to slow down on compatible trials and to speed up on incompatible trials. There was a large faking effect size ($d = 1.14$). However, in most studies the amount of intentional distortion possible on an IAT is less compared to a self-report measure (e.g., Steffens, 2004).

Not all studies show this effect. Banse, Seize, and Zerbes (2001) gave a group of participants instructions to fake an extremely positive attitude toward homosexuals. They could not deliberately change their IAT scores, whereas they could change their self-reported attitude. Egloff and Schmukle (2002) assigned participants ($N = 40$) to one of two groups: a control and a hypothetical job applicant condition. The experimenters gave participants an IAT of anxiety and a self-report anxiety measure. There was no significant difference between the two conditions on the IAT, but there was for the self-report measure. The effect size for the IAT was small ($d = .16$), whereas the significant effect size for the self-report was large ($r = .63$). Egloff and Schmukle (2002) recommended

using larger sample sizes and within-subject designs, a suggestion I heeded in my study below.

In some studies, results related to the fakablity of the IAT are mixed. McDaniel et al. (2009) found that within subject differences between honest ($M = .04$, $SE = .06$) and faking ($M = .39$, $SE = 08$) conditions was significant for an IAT of extraversion ($F(1, 27) = 25.97$, p < .001). There was no difference for an IAT of conscientiousness ($F(1, 27) = 2.05$, p = .163) between honest ($M = .41$, $SE = .09$) and faking ($M = .46$, $SE = .08$) condition.

In sum, past research on the IAT has shown that it has adequate reliability and validity. However, the IAT procedure is not without its critics. Fazio and Olson (2003) say that research on implicit measurement has been largely atheoretical. Other researchers question whether the IAT is a unidimensional measure, because there are two categories in the procedure that are essentially opposites (Blanton, Jaccard, Christine, & Gonzales, 2007). In addition, researchers question whether the IAT score reflects an individual's evaluative association or environmental and cultural associations that an individual has been exposed to in his or her lifetime (Karpinski & Hilton, 2001). Because the research on this type of implicit measure is only nine years old, debate concerning these issues continues. I hope to further our understanding of the IATs with the current study.

The Current Study

For my study, I developed an IAT that measures conscientiousness. I chose the IAT as my implicit measure because of its popularity in the literature. In addition, IAT measures may not be susceptible to large faking effects (e.g., Steffens, 2004), and they do

not introduce a cognitive ability component into the test (Cai, Greenwald, & McFarland, 2004). This is important because other methods and measures to control applicant faking may introduce a cognitive ability component into personality testing (Vasilopoulous & Cucina, 2006). One of the purposes of personality testing in selection is to reduce adverse impact caused by cognitive ability measures. Thus, the IAT is a likely choice for selection. I chose conscientiousness because it is one of the most valid predictors of job performance (Barrick et al., 2001). My study's sample was student participants, so the performance criterion was college grade point average (GPA).

I tested whether using an implicit measure (the IAT) predicts a performance criterion over and above an explicit measure. Several psychologists encouraged researchers to report the incremental validity of explicit and implicit measures used conjunction in predicting a criterion (Banse & Greenwald, 2007; Poehlman, Uhlmann, Greenwald, & Banaji, 2006). Siers and Christiansen's (2008) study compared IATs and traditional self-report personality measures. They did not find support for their hypothesis that IAT measures of personality have incremental prediction beyond self-report. Siers and Christiansen (2008) had students ($N = 200$) complete self-report and IAT measures of the personality traits of extraversion, conscientiousness, and stability. They also measured participants' supervisor ratings of job performance, cognitive ability, GPA, and peer ratings of personality. The extraversion IAT was the only trait that had significant incremental variance beyond the self-reported measure of that trait in predicting GPA (R^2 = .023, $p < .05$).

However, Siers and Christensen (2008) found that their personality IATs correlated significantly with the self-esteem IAT, which lead me to wonder if their IATs

were really measuring self-esteem although they were attempting to measure personality.

Siers and Christensen's personality IATs used valence associations: "positive" and

"negative" were the target categories. Valence associations are used in self-esteem IATs.

Most studies (e.g., Grumm & Collani, 2007; Steffens, 2004; Steffens & Konig, 2006)

with personality IATs use the categories "self", "other", and words that describe the

personality construct and words that describe the opposite of the personality construct,

for example, conscientiousness and negligence words. In my study, I used

"conscientious" and "negligent" words as targets instead of "positive" and "negative"

like Siers and Christiansen (2008) used in their personality IATs. By using the IAT, I

wanted to know if the participant viewed himself or herself as conscientious compared to

"others", a self-trait association. I was not interested in how positively or negatively the

participant views conscientious or negligence terms.

Researchers that used self-trait associations instead of valence associations found

that their personality self-concept IAT did not correlate strongly with self-esteem IATs

(Grumm & Collani, 2007). Grumm and Collani (2007) concluded that the personality

IATs were indeed measuring the personality construct, not implicit self-esteem. There

were low correlations between the measures of IAT self-esteem and IAT extraversion (r

$= .26, p < 0.05$) and IAT self-esteem and IAT neuroticism ($r = -0.38, p < .01$). The way

the target categories are designed in the IAT for personality self-concept makes a

difference if they are measuring self-esteem or personality. I designed my IAT like

Grumm and Collani (2007) and administered a self-esteem IAT to see if my IAT

conscientiousness measure had a strong relationship with it; I hypothesized that it would

not. I pilot tested my developed IAT of conscientiousness (IAT-C) as part of a construct validation procedure.

Researchers advocated measuring explicit and implicit attitudes (and other constructs) in several contexts in a single study (Smith & Conrey, 2007). Research has shown that implicit measures and explicit measures are context sensitive (e.g, Smith & Conrey, 2007). Siers and Christiansen (2008) had a single condition where participants answered the implicit and explicit measures honestly. My study utilized two conditions: an honest condition and a stimulated job applicant condition. Having two conditions, or contexts, allowed me to determine the fakability of the explicit and implicit measures and evaluate how implicit and explicit personality measures predicted a criterion in varying contexts. In my study, I administered explicit and an implicit measures together in two separate sessions. Hoffman et al. (2005) conducted a meta-analysis and found that the order of implicit and explicit measures did not matter with similar relationships observed when the IAT preceded explicit measures ($\rho = .21$) or when explicit measures preceded the IAT ($\rho = .23$).

Construct validation. Implicit measures do not tap into the same constructs as self-reports (Robinson & Neighbors, 2006). I expected that my IAT-C would not correlate strongly with the explicit personality measure. Previous research has shown that implicit measures of personality have low or nonsignificant correlations with explicit measures of personality (e.g., LeBreton, Barksdale, Robin, & James, 2007). However, previous research with an IAT measuring conscientiousness found it to correlate .22 with a traditional self-report personality measure, NEO-FFI (Steffens & Konig, 2006). Another study found that a conscientiousness IAT correlated .18 with a self-report

consciousness scale (Grumm & Collani, 2007). In addition, a meta-analysis found the mean effect size between explicit measures and the Implicit Association Test to be .24 (Hoffman et al., 2005). I expected there to be a low but significant positive correlation between my implicit measure and an explicit measure of conscientiousness.

Several moderators influence the strength of explicit and implicit consistency (for a review, see Hoffman, Gschwender, Nosek, & Schmitt, 2007). Moderators of explicit and implicit consistency are the conceptual similarity between measures, spontaneity of the verbal responses, self-presentation tendencies, and evaluative strength. Because I developed my IAT-C from conscientiousness terms taken from explicit measures, I expected that conceptual similarity may moderate this relationship. I expected there to be a low but significant positive correlation between my implicit measure and an explicit measure of conscientiousness.

Hypothesis 1a: Implicit personality is positively related to explicit personality.

As part of my construct validation process, I assessed how well implicit measures of conscientiousness correlated with each other. How well two implicit measures correlate with each other depends on the extent to which the two measures are tapping similar processes (Houwer, 2003). For example, implicit measurement of attitude latencies (e.g., Fazio, 1995) tap stimulus compatibility mechanisms, and IATs (e.g., Greenwald et al., 1998) tap response compatibility mechanisms. I did not expect to see a strong relationship between my IAT-C and the conditional reasoning measure of achievement motivation (James, 1998). Although the conditional reasoning is an implicit measure, it measured justification mechanisms compared to the IAT-C that measured

response compatibility mechanisms. I expected my IAT-C to correlate with another IAT of conscientiousness (Grumm & Collani, 2007) because they both tap into response compatibility mechanisms. I expected the IAT-C to have a low correlation with the self-esteem IAT, as found in previous research (Grumm & Collani, 2007).

Hypothesis 1b: Implicit personality measures have a positive relationship with each other.

Criterion-related validity. Recent research has found conscientiousness as measured by a self-report measure to be the biggest predictor of high school and college GPA of all of the Big Five personality traits, with correlations ranging from .20 to .24 (Noftle & Robins, 2007). I asked participants to report their GPA. Previous research has found that self-reported GPA is similar to actual GPA for those with high cognitive ability and high GPA but not similar for those with low cognitive ability and low GPA (Kuncel, Credé, & Thomas, 2005). A meta-analysis found that high correlations exist between the actual GPA and self-reported GPA for high school ($r = .82$, $N = 44,176$) and between the actual GPA and self-reported GPA for college GPA ($r = .90$, $N = 12,089$) (Kuncel et al., 2005).

Hypothesis 2a: Conscientiousness, measured with explicit personality measures, is positively related to grade point average (GPA).

In previous research, the conditional reasoning measure of achievement motivation had a moderate correlation with GPA (James, 1998). To my knowledge, no research has tested conscientiousness measured by an Implicit Association Test in predicting GPA. However, an IAT of achievement motivation has predicted test performance (Brunstein & Schmitt, 2004).

Hypothesis 2b: Conscientiousness, measured with implicit personality

measures, is positively related to GPA.

The fakability of the explicit and implicit measures. As mentioned in this

introduction, faking on self-report personality measures occurs in selection (e.g.,

Birkeland et al., 2007). Previous research has found that giving directions to fake an

explicit measure results in a successful manipulation (e.g., Viswesvaran & Ones, 1999).

My experiment has two conditions: answer honestly to the explicit and implicit measures

and respond as if you are applying for a job (fake). I expected there to be within-subject

differences between honest and fake conditions for the explicit measure. I do not expect

to see the same effect for the implicit measure.

Hypothesis 3: There will be differences in fakability between the explicit

measure of conscientiousness and the implicit measure of conscientiousness.

Explicit and implicit measures predict behavioral criteria in three ways: an

additive model, a double dissociation model, and a multiplicative model (Schnabel,

Asendorph, & Greenwald, 2007). In the additive model, explicit and implicit measures

explain different portions of variance in a construct. Double dissociation occurs when an

explicit measure predicts one criterion better than an implicit measure and an implicit

measure predicts a different criterion better than an explicit measure. Researchers found

this double dissociation effect where the implicit self-concept of being shy predicted

observed spontaneous shy behavior better than a self-report measure of shyness. The self-

report measure predicted controlled shy behavior better than implicit shyness

(Asendorph, Banse, & Mucke, 2002). According to the double dissociation model,

Steffens and Konig (2006) found that an IAT of the Big Five predicted spontaneous

37

behavior. The last model is the multiplicative, in which explicit and implicit measures interact to predict behavioral criteria.

I assumed an additive model between my explicit and implicit measures because of the conceptual similarity of the measures. The items of my IAT-C were taken from an explicit measure of conscientiousness. In the additive model, explicit and implicit measures account for variance in predicting a criterion. I hypothesized that implicit measures have incremental variance in predicting GPA above and beyond explicit measures because they are assessing the trait of conscientiousness outside of the respondent's awareness. My hypothesis was that the IAT has incremental variance over the explicit measure in predicting GPA, in both the honest and simulated job applicant (faking) conditions.

Hypothesis 4: Explicit and implicit conscientiousness predict GPA, and implicit conscientiousness has incremental variance over explicit conscientiousness in predicting GPA.

II. METHOD

Participants

The participants were from a large Midwestern university recruited from introductory psychology classes. The subject pool was approximately 66% female and 85% Caucasian students. Upon completion, participants received research credit points for their introductory psychology class.

I conducted power analyses to see how many participants were needed in the pilot study and in the experiment. The pilot testing procedure required 84 participants to achieve a .80 power with a large effect size ($r = .30$) at $\alpha = .05$. The experiment required 170 participants to achieve .78 power with a large effect size ($r = .30$) at $\alpha = .05$. A sample size of 170 had power of .67 to detect the effect of R squared change of .10 in the regression analysis.

I excluded 53 subjects from analysis because either they failed to attend Session 2 or they did not complete most of the surveys. A final total of $N = 442$ participated in this study. There were 172 males and 272 females. The average age was 19 years old, and 80.1% of participants were college freshman. Table 1 displays the demographic information of the experiment participants.

Measures

International Personality Item Pool conscientiousness measure. Goldberg's International Personality Item Pool scales have high correlations with two other Big Five personality measures, the NEO-PI-R (McCrae & Costa, 1992) and the Hogan Personality

Inventory (Hogan, 1992) (Goldberg, 1999). This study used Goldberg's 20-item scale of conscientiousness (see Appendix A). The response format is a five-point graphic rating scale: 1 = *Very Inaccurate,* 2 = *Moderately Inaccurate,* 3 = *Neither Inaccurate or Accurate,* 4 = *Moderately Accurate,* and 5 = *Very Accurate.* The International Personality Item Pool conscientiousness measure (IPIP-C) has an established alpha reliability of .88.

Conditional Reasoning Test of Achievement Motivation. The Conditional Reasoning Test of Achievement Motivation is a scale composed of 15 items (James, 1998). Each item consists of a paragraph of information followed by a question that asks the participant for the most logical explanation for the given information (see Appendix B). I instructed respondents to choose one answer from four options: one achievement-oriented answer, one non-achievement-oriented answer (fear of failure), or two distracter answers. Participants received one point for an achievement-oriented answer, zero points for the distracter answers, and a negative one point for a non-achievement-oriented answer.

Respondents' reasoning is based on implicit assumptions or justification mechanisms about what constitutes rational behavior to them. High scores on this measure indicate higher levels of achievement orientation. The Conditional Reasoning Test of Achievement Motivation is a measure of implicit cognitions and has an established alpha reliability of .62 (Bing et al., 2007). It does not correlate with explicit measures of conscientiousness (Filipkowski & Miller, 2008).

Implicit Association Test of conscientiousness. My IAT procedure is similar to Grumm and Collani's (2007) procedure, but my conscientiousness and negligence words were taken directly from the IPIP (see Appendix C). The IAT-C consists of

conscientiousness and negligence attribute words and the participant classifies each word into self and other target categories. The IAT procedure was comprised of seven blocks with a pause between each trial for instructions (Grumm & Collani, 2007).

Participants categorized a stimulus found on the center of the screen to one of the categories on the upper left or right corners of the screen. Participants pressed the "A" key to associate the stimulus word into the left corner. Participants pressed the "L" key to associate the stimulus word into the right corner. If the participant made a mistake, an "X" appeared on the screen and the participant had to correct his or her mistake. Having the participant correct the wrong response is recommended in research; studies utilize this procedure most often and it prevents the participant from rushing through the test by randomly pressing right or left keys (Schnabel, Asendorpf, & Greenwald, 2007).

All the IATs in this study were scored according to the *D* measure, which is an intraindividual reaction time algorithm (Greenwald, Nosek, & Banaji, 2003). There are seven steps to score an IAT measure. First, trials greater than 10,000 milliseconds are deleted. Subjects are deleted if more than 10% of trials have latencies less than 300 milliseconds. Next, "inclusive" standard deviations are calculated for all trials in blocks 3 and 6 and all trials in blocks 4 and 7. Then the mean latency for responses for blocks 3, 4, 6, and 7 are calculated. Researchers then compute two mean difference scores. The mean of block 6 is subtracted from the mean of block 3 and the mean of block 7 is subtracted from the mean of block 4. Each difference score is then divided by its associated "inclusive" standard deviation. Lastly, the resulting *D* score is the equal-weight average of the two resulting ratios.

Implicit Association Test of conscientiousness developed by another

researcher. Grumm and Collani (2007) created an IAT for conscientiousness (IAT-Grumm; see Appendix C). The IAT-Grumm consists of conscientiousness and negligence attribute words and the participant classifies each word into self and other target categories. I administered it to test the convergent validity of my IAT-C. Their IAT had a Spearman-Brown adjusted split-half correlation of .82. It correlated .18 with NEO-FFI-30 measuring conscientiousness (Grumm & Collani, 2007). The number of trials was half compared to my IAT-C. Blocks 1, 2, 3, 5, and 6 each had 20 trials. Blocks 4 and 7 each had 40 trials. Nosek (2005) found that the number of stimulus words for an IAT does not have an effect on the magnitude of the effect, reliability, or relationship with self-report explicit measures. This IAT was scored the same way as the IAT-C.

Self-esteem Implicit Association Test. Grumm and Collani (2007) developed a Self-esteem IAT with a split-half reliability of .81. The Self-esteem IAT consists of positive and negative attribute words and the participant classifies each word into self and other target categories (see Appendix C). Blocks 1, 2, 3, 5, and 6 each had 20 trials. Blocks 4 and 7 each had 30 trials. This IAT was scored the same way as the IAT-C.

Social Desirability Measure. To measure social desirability, I used the impression management and the self-deception subscales found in the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984, see Appendix D). The BIDR is based on two factors: impression management, which is the conscious aspect of social desirability and self-deceptive enhancement, which is the unconscious aspect. The established alpha for the self-deception subscale ranges from .68 to .80, and the alpha for

the impression management subscale ranges from .75 to .86. When the subscales are combined, this measure of social desirability has an alpha of .83 (Paulhus, 1988).

Participants rate the 40 items on a seven-point scale. After reverse coding the negatively keyed items, one point is added for each an extreme response of 6 or 7. All items are summed to give an overall score measure, or each of the two subsection scale scores can be summed to obtain scores for impression management and self-deceptive enhancement.

Demographics. The items included questions regarding age, gender, years of education, college major, college GPA on a 4.0 scale, high school GPA, and ACT scores (see Appendix E). The researchers checked the participants' reported GPA and ACT scores with the records provided by the University. I transformed the ACT scores into percentile rankings and used that as the ACT variable.

Manipulation check survey. The manipulation check for the faking and honest condition consists of a survey given at the end of the experiment (see Appendix F). The survey assessed if participants understood the directions, distorted their answers to appear more favorable, responded as honestly as they could (if applicable), and if they responded and felt like they were in a job applicant position (if applicable).

Procedure

Participants signed up for both conditions of the study at one time. I separated the honest and applicant conditions by one week. Research assistants administered the measures on laboratory computers. First, participants took the demographic questionnaire, Conditional Reasoning Measure of Achievement Motivation, the IPIP Conscientiousness measures, and the Balanced Inventory of Socially Desirable

43

Responding. Then a practice IAT was given before the three IATs. The practice IAT measured a participants preference for flowers or insects. It was used to get the participants familiar with the IAT procedure. The order of the personality IATs was IAT-Grumm, the IAT self-esteem, and then IAT-C. The IATs were created and ran on F-IAT, a free software program.

In the first condition, the experimenter instructed the participants to answer the questions honestly. A week later, participants returned and were given the same measures in the same order but with instructions to respond as if they were applying for a job. Experimenters repeated these instructions before every measure. The condition type, honest and faking is required to determine if faking occurs. A measure was considered fakable if the mean scores for the honest and applicant (faking) condition were significantly different, and the applicant condition mean was higher than the honest condition mean.

Pilot Study

The pilot testing was a construct validation procedure for my developed IAT of conscientiousness (IAT-C). We expected to see a significant relationship between: 1) implicit and explicit measures of personality 2) the two IATs measuring conscientiousness 3) the explicit and implicit measures conscientiousness and GPA. We also expected the implicit measures not to be prone to social desirability concerns. They will not have a significant relationship with measures of social desirability.

Participants. Participants were college undergraduates form a large Midwestern university. A total of $N = 84$ participated in this study. There were 31 males and 53

females. The average age was 19 years old, and 75.3% of participants were college freshman.

Procedure. The procedure was identical to the experiment procedure, except that there was only one condition. Participants completed all of the measures with directions to respond honestly. There was no hypothetical job applicant condition.

Results and Discussion. Six participants were thrown out from each of the three IATs because they failed to meet the selection criteria for inclusion (Greenwald, Nosek, & Banaji, 2003). Table 2 displays the pilot study descriptive statistics. The Cronbach's internal consistency alphas for the IPIP conscientiousness and Balanced Inventory of Socially Desirable Responding were consistent with previous research. The Conditional Reasoning Measure of Achievements Motivation had a lower reliability of .32 when past research has shown alpha levels of .62.

The standard procedure for calculating IAT reliabilities was used (Schnabel, Asendorph, & Greenwald, 2007). Split-half reliabilities of the difference scores of blocks 6 and 3 and blocks 7 and 4 were computed. Blocks 3 and 4 are combined discrimination tasks and blocks 6 and 7 are reverse combined discrimination tasks. The IATs had low reliabilities (IAT-C r_{tt} = .29, p <.01; IAT-Grumm r_{tt} = .57, p < .01; IAT Self-esteem r_{tt} = .20, *ns*). I calculated the Cronbach's internal consistency alpha as well. These were higher (IAT-C r = .77; IAT-Grumm r = .86; IAT Self-esteem r = .67). Higher alphas compared to split-half reliabilities have been mentioned in the literature on the IAT (Schnabel, Asendorph, & Greenwald, 2007).

Table 2 displays the correlations among the pilot study variables. Results found that the explicit self-report measure, the IPIP conscientiousness, was fakable. It correlated

with the self-deceptive enhancement subscale of the Balanced Inventory of Socially Desirable Responding ($r = .43, p < .01$) and with the impression management subscale ($r = .53, p < .01$). The implicit measures, the Conditional Reasoning Measure of Achievement Motivation and the two IATs of conscientiousness, did not have significant correlations with the Balanced Inventory of Socially Desirable Responding subscales; these measures were not prone to social desirability, or faking, concerns. The Conditional Reasoning Measure of Achievement Motivation was the only implicit measure that correlated with academic performance; it correlated with GPA ($r = .24, p < .05$) and with the ACT ($r = .38, p < .01$). The IPIP conscientiousness correlated with high school GPA ($r = .25, p < .05$).

In sum, the explicit personality measures were fakable compared to the implicit. The Conditional Reasoning Measure of Achievement Motivation and IPIP conscientiousness predicted academic performance, although the IATs did not. A notable finding is that my developed IAT of conscientiousness had a significant, moderate correlation with another developed IAT of conscientiousness ($r = .30, p < .01$). I continued with the experiment because we found adequate construct validity for our developed IAT conscientiousness measures. I tested my hypotheses with a larger sample size in the experiment.

III. RESULTS

Table 3 displays the results of the manipulation check, which was used to determine if the participants understood and followed the instructions in both conditions. In the honest condition 98.6% (N = 428) of participants reported they responded honestly. In the applicant condition, 93.7% (N = 414) of participants reported they responded as if they were applying for a job. In both conditions, the majority of participants reported they understood the directions: 98.2% (N = 434) in honest and 99.1% (438) in applicant condition. In the honest condition, 10.4% (N = 46) of participants reported they distorted their answers to appear more favorable, whereas 23.8% (N = 105) in the applicant condition said they did.

According to the IAT scoring procedure, a participant's data is thrown out if 10% of a subject's trials have response latencies less than 300 milliseconds (Greenwald, Nosek, & Banaji, 2003). In the honest condition, 5.4% (N = 24) of IAT-Grumm, 5.4% (N = 24) of the esteem IAT, and 7.91% (N = 35) of the IAT-C data points were discarded. In the applicant condition, 10.6% (N = 47) of IAT-Grumm, 11.3% (N = 50) of the esteem IAT, and 16.9% (N = 75) of the IAT-C data points were discarded. The participants with discarded IAT data were included in further analyses, they were just missing an IAT score.

Tests of Hypotheses.

      Hypothesis 1a. Table 4 displays the correlation matrix of the experiment variables. Results partially supported Hypothesis 1a; implicit personality has a relationship with explicit personality. The IAT-C correlated with the IPIP conscientiousness in the honest ($r = .11, p < .05$) and applicant ($r = .18, p < .01$) conditions. The applicant condition IAT-C did not have a relationship with the IPIP conscientiousness. The honest condition Conditional Reasoning Measure of Achievement Motivation had a low, negative relationship ($r = -.13, p < .01$) with the honest IPIP conscientiousness. The applicant condition Conditional Reasoning Measure of Achievement Motivation and IPIP conscientiousness did not have a significant relationship.

      Hypothesis 1b. The results partially supported Hypothesis 1b. The results gave moderate support for the construct validity of my developed IAT of conscientiousness (IAT-C). The IAT-C had a significant relationship with another IAT of conscientiousness (IAT-Grumm; Grumm & Collani, 2007) in the honest condition ($r = .39, p < .01$) and in the applicant condition ($r = .43, p < .01$).

      The conditional reasoning measure of achievement motivation did not have a relationship with the IATs measuring conscientiousness. The two IATs measuring conscientiousness had a relationship with the IAT measuring self-esteem. In the honest condition, the IAT-C had a significant relationship with implicit self-esteem ($r = .39, p < .01$), as did the IAT-Grumm with self-esteem ($r = .37, p < .01$). In the applicant condition, the IAT-C had a significant relationship with implicit self-esteem ($r = .25, p < .01$), as did the IAT-Grumm with self-esteem ($r = .37, p < .01$).

Hypothesis 2a. I predicted that conscientiousness, measured with explicit personality measures, has a positive relationship with grade point average (GPA). The results supported Hypothesis 2a. The self-report IPIP conscientiousness correlated with academic performance in the honest ($r = .12$, $p < .05$) and applicant ($r = .11$, $p < .05$) condition.

Hypothesis 2b. I predicted that conscientiousness, measured with implicit personality measures, has a positive relationship with GPA. The results partially support Hypothesis 2b. The IATs measuring conscientious did not have a significant relationship with GPA. The Conditional Reasoning Measure of Achievement Motivation was the only implicit measure that correlated with academic performance. The measure correlated with college GPA in the honest condition ($r = .23$, $p < .01$) and in the applicant condition ($r = .26$, $p < .01$).

Hypothesis 3. I predicted that there will be differences in fakability between the explicit measure of conscientiousness and the implicit measure of conscientiousness The results support Hypothesis 3; the explicit measure of conscientiousness is fakable. Paired-samples t-tests showed that participants were able to raise their score on the IPIP conscientiousness. Table 5 shows the results of the t-tests. Participants were unable to fake the IATs of conscientiousness and actually lowered their score in the applicant condition. The Cohen's *d* for the measures showed a moderate effect size.

Hypothesis 4. I predicted that explicit and implicit conscientiousness predict GPA, and implicit conscientiousness has incremental variance over explicit conscientiousness in predicting GPA. The results did not support Hypothesis 4. I tested Hypothesis 4 with a hierarchal regression analysis that was performed for both conditions

to assess whether the implicit measures predicted GPA over and above the explicit

measure. Table 6 displays the results of this analysis. There was not a significant R^2

change when the implicit measure was added to the regression analysis; it did not add

incremental variance over the explicit measure.

IV. DISCUSSION

Summary

Consistent with previous research, explicit personality measures were fakable compared to the implicit personality measures. Results from the pilot study showed that explicit personality measures correlated with social desirability measures whereas the implicit measures did not. In the experiment, participants were able to increase their conscientiousness scores on the self-report measure in the applicant condition. However, the scores from the IATs and the conditional reasoning measure decreased in the applicant condition. Participants were unable to successfully fake these measures to be hired for the hypothetical job. Some researchers believe that social desirability measures are not perfect proxies of faking behavior. Thus, this study shows that implicit measures of personality are not as susceptible to faking with the use of social desirability measures and a within-subjects design with honest and applicant conditions.

The results of my manipulation check showed that instructions to fake a personality measure is an effective induction. Most of the participants revealed that they understood the directions and responded accordingly. In the honest condition ten percent of respondents said they distorted answers to appear favorable. This percentage rose to twenty-four in the applicant condition. It is unclear from the results of my study what strategies participants used to appear more favorable. Researchers have proposed several moderators of the intention to fake to actual faking behavior, such as responder's knowledge of the measured construct, item transparency, and opportunity (McFarland &

Ryan, 2006). Participants were able to change their scores on the implicit measures to a moderate degree, although this effect was in the negative direction. They did not understand what strategies were required to raise their conscientiousness score on the implicit measures.

A notable finding is that my developed IAT of conscientiousness had a significant, moderate correlation with another developed IAT (Grumm & Collani, 2007). My IAT utilized short phrases as well as single target words, which is unlike from previous IATs. Future research may include phrases as part of the IAT procedure.

The conscientiousness IATs in my study, had significant moderate correlations with the IAT of self-esteem. This is problematic because I wanted to measure personality self-concept and not self-esteem. Previous research had the same results (Siers & Christensen, 2008; Grumm & Collani, 2007). Variance attributable to the measurement method (IAT) rather than the construct of interest (Conscientiousness) is common method variance (Campbell & Fiske, 1959), and calls into the validity of the results. We cannot conclude that the Conscientiousness IAT measures were a strong measure of the construct of Conscientiousness (at least as measured by traditional self-report measures) as they exhibited higher correlations with IAT self-esteem measures. One interpretation of this finding is that IAT Conscientiousness measures are not measuring Conscientiousness, but tapping the construct of self-esteem, or merely correlated due to common method variance such as overall reaction time. One may argue that the construct that the IATs are measuring (apparently some form of the construct on self-esteem) is important, but not that the IATs can serve as a substitute for self-report Conscientiousness measures. Although the results of the study were not what we expected, we have conclusively

answered the question if IATs can be used as a substitute for self-report

Conscientiousness, and our opinion is no.

The conditional reasoning measure and self-report conscientiousness measures predicted academic performance although the IATs did not. My hypothesis was not supported that an IAT of conscientiousness shares variance with GPA. A measure of academic performance, the ACT, only had a significant moderate correlation with GPA. Many other variables can predict GPA but an IAT of conscientiousness is not one of them. The construct of conscientiousness as measured with an explicit measure had a significant, low correlation with GPA. My hypotheses assumed an additive model; explicit and implicit measures account for variance in predicting a criterion. I should have revised my model based on the results that the IAT did not predict the criterion.

Previous research has shown that implicit measures of personality have low or nonsignificant correlations with explicit measures of personality (LeBreton et al., 2007). Because I developed my IAT from conscientiousness words taken from the IPIP, I expected that conceptual similarity would moderate explicit-implicit relationship. I found a low but significant relationship with the explicit and my implicit measure. Even more concerning is that the explicit and implicit measure correspondence was nonsignificant in the applicant condition. Researchers have assumed that context could moderate the correspondence. My study showed that implicit and explicit measures were not measuring conscientiousness in the faking sample. In an applicant setting, this would be a cause for concern. An explanation for this result is that participants do not know how to fake the implicit measures and they are doing so at varying degrees.

Implications

　　If researchers are interested in assessing personality, measuring a trait with both explicit and implicit measures may be necessary to get an accurate reading of an individual's trait level. Explicit measures may tap into self-attributed personality, and implicit measures may tap an unconscious part of personality (Gawronski, Hofman, & Wilbur, 2006; Murray, 1938). The results of this study did not find support for IATs predicting the same outcome as an explicit measure. Researchers have stated that explicit and implicit measures may have different antecedents and may predict outcomes differently (e.g., Bing et al., 2007). If explicit and implicit measures have a small relationship and do not predict the same outcomes when they are supposedly measuring the same construct, the validity of using implicit measures like the IAT in research studies, is called into question.

　　I/O psychology practitioners, HR departments, and managers often administer self-report personality measures in selection. Researchers are concerned with the fakability of these measures. My study showed that an implicit measure, IAT-C, is less fakable than an explicit measure, the IPIP-C. However, practitioners should not use IATs for selection purposes because of concerns with their construct and criterion validity.

Limitations

　　This study has several limitations. First, participants were in a hypothetical job applicant condition and with directions to answer the measures as if they are applying for a job. I have no way of knowing that job applicants fake to the degree that these participants did. Some researchers believe that the fake-good format is a hypothetical

exaggerated condition that is not representative of real-life settings (Smith & Ellingson, 2002).

Second, I used GPA as my performance criterion, which might not be representative of real-world workplace situations. Practitioners use personality measures in selection because certain personality traits predict job performance (e.g., Barrick, Mount, & Judge, 2001). GPA is not job performance. However, a meta-analysis between GPA and job performance found a corrected correlation in the mid .30s (Roth, BeVier, Switzer, & Schippmann, 1996). In addition, personality measures in selection assess traits used to predict performance later. In this study, the criterion of GPA is determined before participants take my measures. Lastly, I used a within-subjects design in my study. Within-subjects designs have greater statistical power than between-subject designs; however, they are sensitive to threats to validity such as history, testing, and maturation (Mesmer-Magnus & Viswesvaran, 2006).

Future Research

There are numerous possibilities for future research in this domain, concerning both applicant faking on personality measures and using implicit measures in applicant selection. Researchers need to test empirically the applicant faking models. In addition, a universal definition of faking behavior needs to be integrated within all future research. Another interesting possibility is to look at differences in applicant faking across cultures. A meta-analysis of various measures of self-enhancement found westerners self-enhanced more than East Asians ($d = .84$; .95 $CI = .67,95$) (Heine & Hamamura, 2007). If self-enhancement is a trait that may influence faking behavior, cultural differences in

self-enhancement may have implications for organizations that are selecting applicants cross-culturally.

Concerning implicit measures, researchers have advocated using polynomial regression to model interactive curvilinear relationships in three-dimensional space to demonstrate the relationship between explicit cognitions, implicit cognitions, and behavioral outcomes (Bing et al., 2007). Future research can examine behaviors when explicit and implicit personality scores do not match. Bing et al. (2007) has demonstrated this for achievement motivation with the conditional reasoning test and a traditional self-report measure. Different behavior patterns may emerge if researchers measure someone as being high in conscientiousness on an implicit measure and low in conscientiousness on an explicit measure and vice versa (assuming faking has not occurred).

Future research can use other implicit personality measures besides the IAT. The Go/No-Go Association Task (GNAT) is different than the IAT because it uses two categories not four, target categories are not paired, and only a single response is required (Boldero, Rawlings, & Haslam, 2007). If the stimulus term belongs to a category the participant presses the key, and if the stimulus term is a distracter the participant does not press the key. Boldero et al. (2007) showed that a GNAT measuring the Big 5 personality traits was reliable and had convergent validity. In addition, researchers have advocated using different types of implicit measurement tools such as the Single-Target IAT, Extrinsic Affective Simon Task, Affects Misattribution Procedure (for a review, see Schnabel, Asendorph, & Greenwald, 2007). Researchers can examine what characteristics of the implicit measure and the testing situation make an implicit measure more susceptible to faking.

Future research can differ the outcomes that IAT and explicit measures predict. Previous research has shown that IAT may be better at predicting spontaneous behavior, whereas explicit measures are better at predicting controlled and planned behavior (e.g., Steffens & Konig, 2006). Steffens and Koning (2006) conducted their study with students and the behaviors were not workplace related. Future research can examine controlled versus spontaneous behavior measuring conscientiousness implicitly and explicitly and predict actual workplace performance criteria, such as organizational citizenship behaviors.

Conclusion

This research was interested in the possibility of an implicit measure serving to complement the ACT in predicting college GPA. Although I were able to replicate IAT methodology and build an IAT measure, it is unclear if it is measuring conscientiousness. The Conditional Reasoning test was a better predictor of college GPA than the explicit conscientiousness measure. The Conditional Reasoning test significantly correlated with the ACT, suggesting it may not complement the ACT well. The undesirable psychometric properties of the IAT measures, coupled with the fact that the ACT was a strong predictor of GPA ($r = .51, p < .01$), suggests that future research with the IAT is needed that focuses on improving the psychometric properties of the IAT such as its criterion validity. In sum, conscientiousness IATs are not susceptible to faking behavior, but they do not have a relationship with a performance criterion or predict above and beyond explicit personality measures. I do not recommend using IATs in conjunction with self-report measures in applicant selection.

Table 1.

Demographic Frequencies for Experiment Study Participants

Category	Number	Percentage
Sex		
Female	272	61.5
Male	170	38.5
Education Level		
Freshman	354	80.1
Sophomore	67	15.2
Junior	12	2.7
Senior	8	1.8
Hours Studying weekly		
Zero hours	3	0.7
1-3 hours	83	18.8
3-5 hours	147	33.3
5-7 hours	107	24.2
7-10 hours	59	13.3
10+ hours	43	9.7
Employment Status		
Non-employed	213	48.2
Part-time	211	47.7
Full-time	18	4.1
Job		
Customer service	151	34.2
Medical field	11	2.5
Skilled trades	15	3.4
Administrator	14	3.2
Teacher	6	1.4
Military	6	1.4

Table 2.

Pilot Study Variables' Descriptive Statistics and Bivariate Correlations

Variable	N	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. CR-AM	84	1.05	4.10	(.34)									
2. IPIP-Con	79	3.31	0.72	-.20	(.92)								
3. IAT-C	79	0.25	0.25	.17	.15	(.77)							
4. IAT-GR	79	0.45	0.38	.04	.23**	.30**	(.86)						
5. IAT-Self esteem	78	0.49	0.26	.12	-.07	.45**	.17	(.67)					
6. BIDR-IM	85	4.94	2.93	-.04	.53**	.07	.01	.06	(.65)				
7. BIDR-SDE	85	6.11	3.45	-.05	.34**	-.06	-.01	.07	.44**	(.69)			
8. College GPA	84	2.41	0.73	.24*	.05	-.03	-.05	-.11	.08	-.11	-		
9. HS GPA	84	3.03	0.50	-.01	.26*	.02	.05	.07	-.05	.04	.40**	-	
10. ACT percentiles	78	46.47	2.93	.39**	-.07	.05	.06	.01	-.13	-.05	.30**	.26*	-

Note. Cronbach's Internal Consistency Estimates are shown in the diagnol. *$p < .05$. **$p < .01$

Conditional Reasoning Measure of Achievement Motivation (CR-AM), IPIP Conscientiousness Scale (IPIP-Con), My Implicit Association Test of Conscientiousness (IAT-C), Grumm and Collani's (2007) Implicit Association Test of Conscientiousness (IAT-GR), Grumm and Collani's (2007) Implicit Association Test of Self-Esteem (IAT-self esteem), Paulhus's (1998) Balanced Inventory of Socially Desirable Responding Impression Management Subscale (BIDR-IM), Paulhus's (1998) Balanced Inventory of Socially Desirable Responding Self-Deceptive Enhancement Subscale (BIDR-SDE).

Table 3.

Experiment Manipulation Check Descriptive Statistics

Category	Number	Percentage
Honest condition		
Distorted answers to appear more favorable	46	10.4
Responded honestly	428	96.8
Understood instructions	434	98.2
Applicant condition		
Distorted answers to appear more favorable	105	23.8
Responded as applying for job	414	93.7
Understood instructions	438	99.1

Table 4.
Correlation Matrix of Study Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. IPIP Con. Honest	(.91)															
2. IPIP Con. Applicant	.46*	(.94)														
3. IAT-C Honest	.11*	.02	-													
4. IAT-C Applicant	.18**	.05	.38**	-												
5. IAT-Gr Honest	.09	.03	.39**	.26**	-											
6. IAT-Gr Applicant	.12*	.06	.44**	.43**	.35**	-										
7. IAT-Est Honest	.06	.01	.39**	.37**	.37**	.34**	-									
8. IAT-Est Applicant	.03	.01	.30**	.25**	.23**	.37**	.32**	-								
9. ACT percentiles	-.04	.02	-.02	.08	-.01	-.01	-.01	.11*	-							
10. College GPA	.12*	.11*	-.03	.07	-.02	.03	-.02	.16**	.51**	-						
11. BIDR- IM Honest	.38**	.22**	.02	.01	.03	-.08	-.01	-.13**	-.08	.06	(.72)					
12. BIDR- IM Applicant	.25**	.41**	-.04	-.02	-.03	-.06	-.01	-.07	.11*	.14**	.57**	(.86)				
13. BIDR-SDE Honest	.36**	.16**	.03	.07	-.08	.02	.00	-.06	-.08	-.09*	.38**	.23**	(.70)			
14. BIDRSDEApplicant	.33**	.32**	.03	.11*	-.04	.09	.02	-.01	-.01	-.03	.37**	.60**	.67**	(.81)		
15. CR-AM Honest	-.13**	.03	-.06	.03	-.02	.02	-.01	.11*	.41**	.22**	-.04	.08	-.03	.06	(.26)	
16. CR-AM Applicant	-.11*	-.01	-.06	.06	-.06	.04	.00	.17**	.44**	.26**	-.05	.06	-.04	.01	.61**	(.30)

Note. Cronbach's Internal Consistency Estimates are shown in diagonal. *p < .05. **p < .01 Conditional Reasoning Measure of Achievement Motivation (CR-AM), IPIP Conscientiousness Scale (IPIP-Con), My Implicit Association Test of Conscientiousness (IAT-C), Grumm and Collani's (2007) Implicit Association Test of Conscientiousness (IAT-GR), Grumm and Collani's (2007) Implicit Association Test of Self-Esteem (IAT-self esteem), Paulhus's (1998) Balanced Inventory of Socially Desirable Responding Impression Management Subscale (BIDR-IM), Paulhus's (1998) Balanced Inventory of Socially Desirable Responding Self-Deceptive Enhancement Subscale (BIDR-SDE).

Table 5.

Differences between Honest and Applicant Conditions

Measure	n	Honest		Applicant		*t*	*d*
		M	SD	M	SD		
IPIP-Con	440	3.55	.65	3.78	.70	-6.79***	.32
IAT-C	350	.29	.23	.24	.22	3.15***	-.16
IAT-Gr	384	.46	.31	.32	.29	8.06***	-.41
IAT-Est	381	.51	.29	.37	.27	8.14***	-.42
BIDR-IM	442	5.37	3.30	6.03	4.59	-3.61***	.17
BIDR-SD	442	6.03	3.47	6.57	4.27	-3.51***	.16
CR-AM	441	1.55	3.93	.94	3.99	3.62***	-.17

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

Table 6.

Hierarchical Regression Analyses of Honest and Applicant Conditions in Predicting College Grade Point Average

	Condition Type			
	Honest		Applicant	
Predictor	ΔR^2	β	ΔR^2	β
Step 1	.26**		.24**	
ACT percent		.51**		.49**
Step 2	.02**		.01**	
ACT percent		.52**		.49**
IPIP Con		.16**		.12*
Step 3	.01		.00	
ACT percent		.52**		.49**
IPIP Con		.17**		.12*
IAT-C		-.06		.00
n	359		326	

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

IV. REFRENCES

Asendorph, J.B. (2007). Editorial: Implicit representations and personality. *International Journal of Psychology, 42*(3), 145-148.

Asendorph, J.B., Banse, R., & Mucke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case for shy behavior. *Journal of Personality and Social Psychology, 83,* 380-393.

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.*

Banse, R., & Greenwald, A.G. (2007). Personality and implicit social cognition research: Past, present, & future. *European Journal of Personality, 21,* 371-382.

Banse, R., Seise, J., & Zerbes, N. (2001) Implicit attitudes toward homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48* (2), 145-160.

Barrick, M.R., Mount, M.K., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do I know and where do I go from here? *Personality and Performance, 9 (*1/2), 9-30.

Barrick, M.R., & Mount, M.K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance, 18* (4), 359-372.

Bing, M.N., LeBreton, J.M., Davison, H.K., Migetz, D.Z., & James, L.R. (2007). Integrating implicit and explicit social cognitions for enhanced personality assessment: A general framework for choosing measurement and statistical methods. *Organizational Research Methods,* 10, 346-389.

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.

Blanton, H., Jaccard, J., Christine, C., & Gonzales, P.M. (2007). Plausible assumptions, questionable assumptions and post hoc rationalizations: Will the real IAT please stand up? *Journal of Experimental Social Psychology, 43,* 399-409.

Boldero, J.M., Rawlings, D., & Haslam, N. (2007). Convergence between GANT-assessed implicit and explicit personality. *European Journal of Personality, 21,* 341-358.

Bott, J.P., O'Connel, 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.

Brunstein, J.C., & Schmitt, C.H. (2004). Assessing individual differences in achievement motivation with the Implicit Association Test. *Journal of Research in Personality, 38*, 536-555.

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.

Cai, N.S., Greenwald, A.G., & McFarland, S.G. (2004). The Implicit Association Test's

    D measure can minimize a cognitive skill confound: Comment on McFarland and

    Crouch (2002). *Social Cognition, 22* (6), 673-684.

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the

    multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

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.

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.

De Houwer, J. (2003). A structural analysis of indirect measures of attitudes. In J. Musch

    & C. Klauer (Eds.) *The psychology of evaluation: Affective processes in cognition*

    and emotion (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum Associates.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of

    implicit measures. In B. Wittenbrink and N. Schwarz (Eds.) *Implicit measures of*

    attitudes (pp. 179-194)*.* New York: The Guildford 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.

Egloff, B., & Schmukle, S.C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology, 83* (6), 1441-1455.

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.

Elligson, 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.

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.

Fazio, R.H., & Olson. M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review Psychology*, *54* (1), 297-327.

Fiedler, 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.

Filipkowski, J.N., & Miller, C.E. (2008, March). An evaluation of the fakability of conscientiousness measures. Paper presented at 29th Industrial Organizational and Organizational Behavior Graduate Student Conference, Denver, CO.

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.

Gawronski, B., Hofmann, W., & Wilbur, C.J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition, 15,* 485-499.

Gawronski, B., LeBel, E.P., & Peters, K.R. (2007). What do implicit measures tell me? *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.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., & Banaji, M.R., (2009). Understanding and using the Implicit Association Test. *Journal of Personality and Social Psychology, 97* (1), 17-41.

Griffith, R.L., Chemielowski, 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.

Grumm, M., & Collani, G. (2007). Measuring Big-Five personality dimensions with the

Implicit Association Test: Implicit personality traits or self-esteem? *Personality*

and Individual Differences, 43, 2205-2217.

Gschwedner, T., Hofman, W., & Schmitt, M. (2008). Convergent and predictive validity

of implicit and explicit anxiety measures as a function of specificity similarity

and content similarity. *European Journal of Psychological Assessment, 24* (4),

254-262.

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.

Heine, S.J., & Hamamura, T. (2007). In search of East Asian self-enhancement.

Personality and Social Psychology Review, 11, 4-27.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-

analysis of the correlation between the Implicit Association Test and explicit self-

report measures. *Personality and Social Psychology Bulletin, 31* (10), 1369-1385.

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.

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. & Evoy, R.A. (2005). Personality inventory faking: A four-dimensional
simulation of dissimulation. *Personality and Individual Differences, 39*, 1307-
1318.

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.

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job
performance. *Psychological Bulletin, 96*, 72-98.

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.

James, L.R. & Mazerolle, M.D. (2002). Personality in Work Organizations. Thousand

    Oaks, CA: Sage Publications, INC.

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.*

    Greenwich: Information Age Publishing.

Karpinski, A., & Hilton, J.L. (2001). Attitudes and the implicit association test. *Journal*

    of Personality and Social Psychology, 81 (50), 774-788.

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.

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).

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The Validity of Self-Reported Grade

    Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of

    the Literature. *Review of Educational Research*, *75*(1), 63-82.

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.

Lane, K.A., Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2007). Understanding and

    using the Implicit Association Test: IV: What I know (so far) about the method.

    In B. Wittenbrink and N. Schwarz (Eds.) *Implicit measures of attitudes* (pp. 59-

    102). New York: The Guildford Press.

LeBreton, J.M., Burgess, J.R.D., & James, L.R. (1997, April) Measurement issues

associated with Conditional Reasoning: Deception and faking. Paper presented at

the Society for Industrial and Organizational Psychology, in Dallas, Texas.

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.

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.

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., Heggestad, 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.

Murray, H.A. (1938). *Explorations in Personality.* New York: Oxford University Press.

Noftle, E. E., & Robins, W. (2007). Personality Predictors of Academic Outcomes. *Journal of Personality and Social Psychology*, *93*(1), 116-130.

Nosek, B.A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology, 40,* 543-550.

Nosek, B.A. (2007). Explicit and implicit relations. *Current Directions in Psychological Science, 16*(2), 65-69.

Nosek, B.A, Banaji, M.R., & Greenwald, A.G. (2002). Math = male, me =female, therefore math ≠ me. *Journal of Personality and Social Psychology, 8* (1), 44-59.

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., 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.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Poehlman, T.A., Uhlmann, E., Greenwald, A.G., & Banaji, M.R. (2006). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Manuscript submitted for publication.

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.

Roth, P.L., BeVier, C.A., Switzer, F.S., & Schippmann, J.S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*, 548-556.

Ryan, A.M., Ployhart, R.E., & Friedel, L.A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83*(2), 298-307.

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.

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.

Schmukle, S.C., Back, M.D., & Egloff, B. (2008). Validity of the five-factor model for the implicit self-concept of personality. *European Journal of Psychological Assessment, 24* (4), 263-272.

Schnabel, K., Asendorph, J.B., & Greenwald, A.G. (2007). "Using Implicit Association Tests for the assessment of implicit personality self-concept." In G.J. Boyle, G. Matthews, and H. Saklofske (Eds.) *Handbook of personality theory and testing.* London: Sage.

Siers, B.P., & Christiansen, N.D. (2008). On the validity of implicit association test measures of personality traits. Paper presented at 23rd Society for Industrial and Organizational Psychology Conference, San Francisco, CA, April 2008.

Smith, E.R., & Conrey, F.R. (2007). "Mental representations are states, not things: Implications for implicit and explicit measurement". In B. Wittenbrink and N. Schwarz (Eds.) *Implicit measures of attitudes* (pp. 247-264). New York: The Guildford Press.

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.

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.

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 familiarly

    and impression management on self-report measure scale scores and response

    latencies. *Journal of Applied Psychology, 85*, 50-64.

Viswesvaran, C., & Ones, D.S. (1999). Meta-analysis of fakability estimates:

    Implications for personality measurement. *Educational and Psychological*

    Measurement, 59, 197-210.

Waters, L.K. (1965). A note on the "fakability"of forced-choice scales. *Personnel*

    Psychology, 18, 187-191.

Zikar, 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.

APPENDIX A

IPIP Conscientiousness Scale (Goldberg, 1999)

On the following page, there are phrases describing people's behaviors. Please use the rating scale below to describe how accurately each statement describes you. Please read each statement carefully, and then fill in the bubble that corresponds to the number on the scale.

Response Options

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

1. Am always prepared.
2. Pay attention to details.
3. Get chores done right away.
4. Like order.
5. Follow a schedule.
6. Am exacting in my work.
7. Leave my belongings around.
8. Make a mess of things.
9. Often forget to put things back in their proper place.
10. Shirk my duties.
11. Do things according to a plan.
12. Continue until everything is perfect.
13. Make plans and stick to them.
14. Love order and regularity.
15. Like to tidy up.
16. Neglect my duties.
17. Waste my time.
18. Do things in a half-way manner.
19. Find it difficult to get down to work
20. Leave a mess in my room.

APPENDIX B

Conditional Reasoning Test of Achievement Motivation (James, 1998)

Sample item:

Going to a job interview is like acting. People who are new to the process are likely to suffer from stage fright and make nervous mistakes. With experience at being interviewed, people learn what questions to expect and what answers impress interviewers. The key to doing well in interviews is to gain this experience at being interviewed.

Which of the following most weakens this argument?
A. Every interviewer has a personal theory as to what the perfect applicant is like.
B. Interviewers usually form first impression of applicants before meeting them based on their resumes or applications.
C. Applicants who are interviewed just after a truly outstanding candidate tend to score less well than they would if interviewed after an average or poor candidate.
D. Interviewers generally prefer short answers to their questions unless they ask the applicant to provide more information on a specific point.

Key: A-achievement motivation, C-fear of failure, B and D distracter items

APPENDIX C

Implicit Association Test Items

IAT-Conscientiousness (IAT-C)

Conscientious: prepared, detail-oriented, likes-order, dependable, diligent, self-disciplined, organized, planful, perfectionist, tidy, and ambitious

Negligent: sloppy, aimless, messy, shirk-duties, forgetful, neglectful, waste-time, procrastinator, careless, disorganized, slacker, and lazy

IAT Self-Esteem (Grumm & Collani, 2007)

Positive words: precious, good, satisfied, pleasant, sunny, happy

Negative words: useless, bad, greedy, unpleasant, unjust, sad

IAT Conscientiousness (Grumm & Collani, 2007)

Conscientiousness: tidy, pedantic, punctual, diligent, dependable

Negligent: sloppy, chaotic, unpunctual, lazy, undependable

APPENDIX D

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

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

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

Self-Deception Subscale
1. My first impressions of people usually turn out to be right.
2. It would be hard for me to break any of my bad habits.
3. I don't care to know what other people really think of me.
4. I have not always been honest with myself.
5. I always know why I like things.
6. When my emotions are aroused, it biases my thinking.
7. Once I've made up my mind, other people can seldom change my opinion.
8. I am not a safe driver when I exceed the speed limit.
9. I am fully in control of my own fate.
10. It's hard for me to shut off a disturbing thought.

11. I never regret my decisions.
12. I sometimes lose out on things because I can't make up my mind soon enough.
13. The reason I vote is because my vote can make a difference.
14. My parents were not always fair when they punished me.
15. I am a completely rational person.
16. I rarely appreciate criticism.
17. I am very confident of my judgments.
18. I have sometimes doubted my intelligence.
19. It's all right with me if some people happen to dislike me.
20. I don't always know the reasons why I do the things I do.

APPENDIX E

Demographics Questionnaire

1. UID #:_____
2. AGE:_____
3. GENDER: male OR female
4. EDUCATION LEVEL (circle one): College Freshman * Sophomore * Junior * Senior
* Graduate Student * other:_____
5. COLLEGE MAJOR:_____
6. How many hours per week do you study or do school work(circle one): zero hours, 1-3
hours, 3-5 hours, 5-7 hours, 7-10 hours, 10+ hours
7. CURRENT EMPLOYMENT STATUS:_____
8. JOB DESCRIPTION:_____
9. YEARS OF EMPLOYMENT (circle one): 0 * 1-2 * 3-4 * 5-6 * 7-8 * other:____
10. GPA in HIGH SCHOOL (on a 4.0 scale):_____
11. GPA in COLLEGE (on a 4.0 scale):_____
12. ACT score:_____
13. SAT score:_____

APPENDIX F

Manipulation Check Survey

Please answer the following questions:

1. Did you understand the instructions of the study?
        Yes or No, Explain:
2. Did you distort your answers on the survey questions to appear more favorable?
        Yes or No, Explain:
3. Did you respond to the questions as honestly as you could? (if applicable)
        Yes or No, Explain:
4. Did you respond to the questions as if you were applying for the position in the job description? (if applicable)
        Yes or No, Explain: