Using the Bogus Knowledge Scale to Detect Individual Differences in Faking:
Examining the Impact of Variance in Applicant Faking

By Jennifer L. Harvel

B.A. in Psychology, May 2005, University of Missouri

A Dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 20, 2012

Dissertation directed by

David P. Costanza
Associate Professor of Organizational Sciences and Psychology

The Columbian College of Arts and Sciences of The George Washington University certifies that Jennifer L. Harvel has passed the Final Examination for the degree of Doctor of Philosophy as of February 2, 2012. This is the final and approved form of the dissertation.

Using the Bogus Knowledge Scale to Detect Individual Differences in Faking: Examining the Impact of Variance in Applicant Faking

Jennifer L. Harvel

Dissertation Research Committee:

David P. Costanza, Associate Professor of Organizational Sciences and Psychology, Dissertation Director

Nils Olsen, Assistant Professor of Organizational Sciences, Committee Member

Dana Dunleavy, Senior Selection Research Associate, Association of American Medical Colleges, Committee Member

Dedication

The author wishes to dedicate this dissertation to her parents, Steve and Cindy Harvel, for

their continued love and support.

Acknowledgments

The author wishes to thank all who made completion of this dissertation possible. First, she would like to thank her dissertation advisor, David Costanza, for his guidance and assistance in developing the main points of the document. She would also like to thank her dissertation committee, Nils Olsen, Dana Dunleavy, Lynn Offermann, and Chuck MacLane, for providing helpful feedback and insight to greatly improve this document. The author would like to thank all of the George Washington Industrial/Organization Psychology program faculty members and students she had the opportunity to work with for sharing their support and technical knowledge. The author would like to further thank her family, friends, and colleagues for their support and encouragement. Their persistent words of motivation and encouragement enabled her to work through and overcome many periods of frustration and apathy. The author would like to especially thank a truly great friend, Rebecca Fraser, for always making herself available to be a sounding board and being the optimistic voice of reason she needed to be able to finish this dissertation.

Abstract of Dissertation

Using the Bogus Knowledge Scale to Detect Individual Differences in Faking:
Examining the Impact of Variance in Applicant Faking

Practitioners and researchers have debated about the extent to which applicant faking impacts the validity and usefulness of non-cognitive measures as well as how response distortion may impact subsequent job performance. One reason for the disagreement may be due to the fact that most previous research has not accounted for individual differences in applicants' faking. Individual differences include both the amount of response distortion applicants engage in, and magnitude of faking, and the extent to which individuals differ in their faking behavior, or variability in faking. The current study used a bogus knowledge scale, which asks individuals to indicate their knowledge of or familiarity with fictitious concepts or items (Anderson, Warner, & Spencer, 1984; Pannone, 1984), to assess differences in faking behavior across individuals. First, the extent to which individual differences in faking impacts the criterion-related validity of non-cognitive measures in predicting task performance in an assessment center (AC) was investigated. Next, applicant AC scores were examined to determine whether there were performance differences among individuals who distort their responses relative to those who respond honestly. As part of the application process for a government-wide internship program, 970 individuals completed a biodata inventory, bogus knowledge scale, and participated in an AC. Results indicated that individual differences in applicant response distortion did not impact the extent of faking on the prediction of task performance in the AC. Further, no differences in performance were observed between applicants who engaged in high faking magnitudes relative to

those who respond honestly. The implications of this study are discussed as well as

directions for future research.

Table of Contents

List of Figures

x

List of Tables

Chapter 1: Introduction

Mary is applying for a job at a company, and as part of the application process she is given a series of personality and other inventories to complete. These assessments are designed to gather job-relevant information that is indicative of future behavior on the job. But what if Mary intentionally provides inaccurate information about herself when completing the inventories? And worse, if Mary is hired on the basis of this inaccurate information, will this impact her ability to perform successfully once on the job?

A number of researchers have attempted to answer these questions. The issue of response distortion – or applicant faking – has become an increasing concern due to the popularity of using self-report non-cognitive measures in personnel selection. Meta-analytic results have shown that personality tests (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Hurtz & Donovan, 2000; Tett, Jackson, & Rothstein, 1991) and biographical data, or "biodata" inventories (Schmidt & Hunter, 1998; Schmidt & Rothstein, 1994), are valid predictors in a variety of jobs. Organizations use these non-cognitive measures because they are relatively inexpensive, easy to administer, and have been typically shown to have less adverse impact than cognitive assessments (Hough, 1998a; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Pulakos & Schmitt, 1996; Schmidt & Hunter, 1998). However, one major limitation associated with these non-cognitive measures is that they are susceptible to applicant faking.

Consider the following items, which are typical biodata items used by organizations (applicants respond using an Agreement scale): "People would say that I'm good at solving problems," "People have told me that I am an effective communicator," "People would say that I am very flexible and can easily adapt." If these statements do

not accurately describe the applicant, this would be important information for organizations to know when making selection decisions because they have implications for a potential employee's ability to perform the job successfully. However, many applicants may be inclined to respond to these types of test items dishonestly when a job opportunity is at stake.

A volume of research has focused on examining the effects of applicant faking on non-cognitive measures. While this research has consistently shown that applicants are able inflate their responses on personality tests (Hough et al., 1990; Viswesvaran & Ones, 1999) and biodata inventories (Lautenschlager, 1994; Levashina, Morgeson, & Campion, 2009; McFarland & Ryan, 2000) when instructed to do so, there is less agreement about the consequences of applicant faking. In fact, there is a contentious debate within the literature regarding the prevalence of applicant faking and its effects on the validity of selection decisions in applied settings.

One group of researchers suggests that faking occurs infrequently and does not negatively affect the prediction of performance (Abrahams, Neumann, & Githens, 1971; Hough, 1998; McCrae & Costa, 1983; Ones, Viswesvaran, & Reiss, 1996). In contrast, another camp of researchers argues that faking alters the psychometric properties of non-cognitive tests and attenuates relationships with performance criteria (Bradley, O'Shea, & Hauenstien, 2002; Douglas, McDaniel, & Snell, 1996; Dunnette, McCartney, Carlson, & Kirchner, 1962; Ellingson, Sackett, & Hough, 1999; Griffin, Hesketh, & Grayson, 2004; Griffith, 1998; Mueller-Hanson, Heggestad, & Thornton, 2003; Pannone, 1984; Rosse, Stecher, Miller, & Levin, 1998; Van Iddekinge, Raymark, & Roth, 2005).

The impact of applicant faking is an important issue for both researchers and practitioners because it has implications for organizations. Specifically, one major concern is the potential impact of applicant faking on selection decisions. If applicants are able to distort their responses so that they rise to the top of scoring distributions, and consequently disrupt the rank order of applicants, the validity and usefulness of the selection tool may be seriously undermined.

In this dissertation, I explore the possibility that one reason for the current conflicting findings about the impact of applicant faking on non-cognitive measures may be due to individual differences in applicant faking. Although it has been acknowledged that individuals vary in the amount that they fake (Mueller-Hanson, et al., 2003; Mueller-Hanson, Heggestad, & Thornton, 2006), individual differences in faking has received little attention in the literature to date (McFarland & Ryan, 2000; Vasilopoulos, Reilly, & Leaman, 2000). For the purposes of this dissertation, individual differences in applicant faking includes both the magnitude (i.e., the average amount that applicants distort their responses) as well as variability (i.e., the extent to which individuals differ in their faking) of response distortion (Komar, Douglas, Komar, & Robie, 2008).

The ability to capture individual differences in response distortion is important for several reasons. If there is minimal variability in the extent to which applicants distort their responses (i.e., everyone fakes to the same extent), the rank-order and predictive validity of applicants are less likely to be affected. However, if individuals do fake to varying magnitudes on selection tests (i.e., some people slightly inflate their responses and others severely distort their responses), individuals who engage in extreme faking are likely to be over-represented at the top of distribution lists. As a result, the criterion-

related validity is more likely to be adversely affected when there is a high degree of variance in applicant faking (e.g., varying levels of faking magnitude and variability). This may help explain why certain studies have shown faking to have negative consequences (e.g., Douglas et al., 1996; Dunnette et al., 1962; Holden & Jackson, 1981; Mueller-Hanson et al., 2003; Pannone, 1984; Rosse et al., 1998) and others have not (e.g., Abrahams et al., 1971; Hough, 1998; McCrae & Costa, 1983; Ones et al., 1996). Therefore, one goal of this study is to examine the impact of individual differences in response distortion on the predictive validity of non-cognitive measures.

If individual differences in applicant faking do impact the criterion-related validity, another critical issue to consider is the implication that this has for organizations relying on non-cognitive measures (i.e., biodata inventories) to make selection decisions. The long-term consequences for organizations could be dire if dishonest and unqualified individuals are faking their way to the tops of selection lists and being hired over qualified applicants who respond honestly. If organizations are hiring these individuals, this could affect a variety of job-related outcomes (e.g., job turnover) and, more specifically, job performance. In other words, applicants who feel a greater incentive to present themselves in a more positive light by inflating their responses would be more likely to be hired, without necessarily being able to perform successfully on the job (Mueller-Hanson et al., 2003).

In this dissertation, I will test the hypothesis that individuals who engage in high magnitudes of response distortion to obtain employment will also have lower task performance relative to those who respond honestly, as measured by performance in an assessment center (AC). Support for this hypothesis will support previous conjectures

4

that faking is a cause for concern for organizations. Therefore, this is an issue that warrants further review in order to better understand the impact of applicant faking on non-cognitive measures as well as whether faking has implications for organizations in terms of task performance.

The present research suggests two reasons as to why faking is expected to have negative implications for task performance. First, it is possible that applicants who severely distort their responses to get a job do not have the skills and qualities needed to perform the required job tasks as well as honest respondents (which is why they inflate their responses). Individuals who inflate their responses for this reason are at the top of distribution lists because they engaged in extreme faking behavior (high faking magnitude) and will likely have test scores that include irrelevant faking variance, which is less likely to be meaningfully predictive of task performance. Thus, applicants who distort their responses because they do not have the knowledge or skills needed to perform the job would likely translate in poorer task performance. Further, the more an applicant misrepresents himself or herself, it is assumed that there will be a greater disparity between the individual's expected and actual performance. Because this disparity should be the largest among applicants who engage in high magnitudes of faking, the associated performance for these individuals is expected to be lower than for those with low faking magnitudes (i.e., those who respond honestly).

Second, it is possible that the personality characteristics that lead individuals to engage in extreme response distortion may also be related to poor task performance. For instance, people who distort their responses in a selection context may have certain traits that are associated with poor task performance (e.g., low conscientiousness, high

neuroticism). Therefore, it is possible that the applicants who severely misrepresent themselves to get a job may not perform the job tasks as well as applicants who obtain high scores by responding honestly because their personality traits are related to poor task performance.

Thus, it is useful to identify which individuals are faking in order to determine whether they will be able to perform successfully on the job once hired. If the conjecture that individuals who engage in high magnitudes of faking have poorer task performance (as measured by AC scores) is supported, this will provide support for the notion that faking is a cause for concern for organizations that may be more likely to select these individuals. Since the individuals who engage in extreme faking behaviors are more likely to be at the top of the predictor score distribution, these individuals are most likely to be hired first, which could have dire consequences for organizations. Investigating this relationship contributes to existing literature by directly examining the potential impact of individual differences in faking on predicting job performance, and seeks to provide insight about the extent to which faking may pose a threat to subsequent task performance.

In order to answer the question of how variance in applicant faking impacts the validity of non-cognitive measures and subsequent job performance, a more precise measure of applicant faking is needed. One promising approach that has received little attention to date is the bogus knowledge scale (Anderson et al., 1984; Pannone, 1984; Levashina et al., 2009). The bogus knowledge scale asks individuals to indicate their knowledge of or familiarity with fictitious concepts or items. This approach has several advantages over existing social desirability scales, including the ability to detect untrue

6

responses. Specifically, the bogus knowledge scale is able to gauge how frequently and to what extent (the faking magnitude) an individual distorts his or her responses. Using this type of measure to detect response distortion may provide a better understanding of the impact of applicant faking on the validity of non-cognitive measures and subsequent task performance.

In summary, the primary goals of the current study are to: 1) use a bogus knowledge scale to examine the extent to which individual differences in faking impacts the criterion-related validity of non-cognitive measures in predicting task performance in an AC, and 2) to determine whether there are differences in applicant AC task performance scores among individuals who distort their responses relative to those who respond honestly. This study is relevant in light of the increased use of personality inventories as well as the research suggesting that faking on non-cognitive measures may have consequences for the accuracy and fairness of selection decisions (Douglas et al., 1996; Griffith, 1998; Mueller-Hanson et al., 2003; Rosse et al., 1998).

Specifically, establishing a link between faking and task performance would provide support for using the bogus knowledge scale as a useful screening tool and may provide guidance in decisions about how to address applicants who engage in extreme faking magnitudes (e.g., removing them from an applicant pool). Using the bogus knowledge scale to identify applicants who should be removed from the selection process may result in a selection process that is fairer for the applicants who do not engage in response distortion. Additionally, using a bogus knowledge scale early in the selection process may be cost-effective for organizations, as individuals who are removed from the applicant pool based on amount of bogus knowledge endorsement would not participate

in subsequent assessments that may be more costly. Thus, the current study is useful to researchers and practitioners because it seeks to provide further insight about the extent to which faking poses a threat to selection processes utilizing non-cognitive measures as well as subsequent performance. Results may provide guidance as to how to address individuals who engage in extreme magnitudes of faking during the selection process.

Chapter 2: Background and Hypotheses

Applicant Faking on Non-Cognitive Measures

In the past decades, applicant faking on non-cognitive measures has received extensive attention within the literature. This research has equivocally shown that applicants are able inflate their responses on self-report personality and biodata measures when instructed to do so (Hough et al., 1990; Lautenschlager, 1994; Levashina et al., 2009; McFarland & Ryan, 2000; Viswesvaran & Ones, 1999). Additionally, there is evidence suggesting that as many as half of an applicant sample will engage in some level of response distortion (Donovan, Dwight & Hurtz, 2003; Fluckinger, McDaniel, & Whetzel, 2008; Griffith, Chmielowski, & Yoshita, 2007). This research has caused researchers and practitioners to become increasingly concerned about the potential negative consequences associated with applicant faking. The present research suggests that one reason why identifying extreme fakers may be advantageous is the possibility that applicants who convey false information about themselves to get a job may not have the necessary skills to perform the job tasks successfully or may have personality traits associated with poor task performance. Thus, it is useful to identify which individuals are faking in order to understand whether they may be more likely to be unsuccessful on the job if hired. A variety of approaches of assessing faking have been used in conjunction with non-cognitive measures. The next section contains a review of one of the most common approaches to detecting applicant faking and discusses why an improved measure is needed.

Social Desirability Scales

One of the most common approaches to detecting applicant faking on self-report measures has been the use of social desirability scales (Marlowe-Crowne scale; Crowne & Marlowe, 1964, Edwards Social Desirability scale; Edwards, 1957; Unlikely Virtues; Hough, 1998; Assessment of Background and Life Experiences (ABLE) Validity Scale; Hough et al., 1990; and the Balanced Inventory of Desirable Responding or BIDR, Paulhus, 1998). Social desirability scales are designed to assess a response strategy (i.e., socially desirable responding) whereby respondents attempt to convey an overly positive image of themselves (Paulhus, 1998). Specifically, these scales are designed to assess the extent to which individuals exaggerate positive traits and minimize negative qualities in order to provide an overly positive impression of themselves. These scales ask respondents to self-report how frequently they engage in common, but generally undesirable behaviors (e.g., cursing, littering, etc.). Individuals are identified as responding honestly or engaging in response distortion based on how frequently they endorse these "unlikely virtues." This is based on the assumption that people engage in these behaviors, and when they claim they do not, they are being dishonest. For instance, individuals who claim that they never curse or litter are assumed to be distorting their responses.

Social desirability scales are based on Paulhus' (Paulhus, 1984, 1986, 1991) conceptualization of social desirability. This model distinguishes two components of social desirability: self-deception and impression management. Self-deception refers to an individual's tendency to respond in a favorable way that the individual believes is truthful. This type of responding is usually done unconsciously, and is believed to be consistent across situations (Paulhus et al., 1995). Items assessing self-deception are

typically geared toward identifying individuals who truly believe their positively inflated response (i.e., "My first impression always turns out to be right"). Researchers are less concerned with self-deception because such responding is not intentional and is believed to be a part of the individual's personality (Paulhus, 1991).

Impression management (IM), on the other hand, is described as a person's tendency to intentionally present false information to create a positive impression to others (Paulhus, 1991, 2001). Since the goal of IM is to be perceived favorably by a particular audience, responses are likely to change across situations (Paulhus, Bruce, & Trapnell, 1995). Items assessing IM are targeted at identifying individuals who are purposefully tailoring responses to create the most positive self-presentation (i.e., "I always drive the speed limit"). Researchers and practitioners believe that when applicants intentionally select the answer on a selection measure that they believe will make them look the most favorable instead of the most accurate answer, they are engaging in IM (Paulhus & Reid, 1991). Therefore, IM has been the focus for much of the faking research because it is assumed to be applicants' intentional misrepresentation of themselves.

Despite the popularity and prevalence of social desirability scales, research has started accumulating suggesting that these scales are inadequate for assessing faking (Christiansen, 1998; Ellingson et al., 1999; Griffith, Rutkowski, Gujar, Yoshita, & Steelman, 2005; McFarland & Ryan, 2001; Snell, Sydell, & Lueke, 1999; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001; Zickar & Drasgow, 1996). One criticism of social desirability scales is that they capture true personality variance (Cunningham, Wong, & Barbee, 1994). Research has repeatedly shown that social desirability scales

correlate with a number of personality characteristics, such as adjustment, conscientiousness, emotional stability, and integrity (e.g., Borkenau & Ostendorf, 1989; Cunningham et al., 1994; Helmes & Holden, 1986; Hurd, Barrett, Miguel, Tan, & Lueke, 2001; Ones et al., 1996; Zerbe & Paulhus, 1987). In other words, individuals who are high in conscientiousness and integrity may not engage in these unlikely virtues, but their honest responses are likely to result in high social desirability scores and being flagged for distorting their responses. This is concerning for researchers because it will result in an unrealistically high false-positive rate. Furthermore, this research warns that attempts to correct for faking by adjusting scores may reduce relevant trait variance, which can lower criterion-related validity (Hogan & Nicholson, 1988; Hogan, Hogan, & Roberts, 1996).

Another major weakness of social desirability scales is that they use a summed scale-score to assess response distortion; thus, the nature of the scale is unable to determine when an applicant distorts a response to a specific item. Further, social desirability scales are unable to differentiate between varying amounts of response distortion, and cannot detect when an applicant provides an outright untruthful response. For instance, one applicant with a true score of a 4.5 on a particular item may slightly exaggerate his or her response by endorsing a "6" on a seven-point scale. Another applicant who has a true score of a 1 on the same item may blatantly lie by endorsing a "6" on a seven-point scale. In this example, the first applicant is exaggerating their response to make a positive impression in a subtle way by marginally distorting his or her response along the scale, whereas the second applicant is blatantly lying by distorting his or her response several points from where it should be.

The first instance is more in line with Paulhus' (1991) notion of self-deception in that the individual slightly exaggerates his or her response to convey a positive image. This type of response distortion is expected to be less problematic because it is relatively close to the individual's "true score" on the trait. On the other hand, the second instance is indicative of IM, in which the individual exaggerates his or her response to an extent that is inaccurate. This distinction is important because these are two examples of ways individuals may intentionally alter responses that are fundamentally different, though social desirability scales are unable to distinguish between the two behaviors.

The inability to detect variance in applicant faking is particularly important in light of research suggesting that faking is not consistent across applicants (McFarland & Ryan, 2000; Mueller-Hanson et al., 2006). As the previous example illustrated, there is an important conceptual difference between slight exaggeration and blatant fabrication. Making this distinction is important because it may have implications for how these individuals behave on the job in the future. Specifically, Paulhus (2001) suggested that attempts to put one's best foot forward by embellishing one's existing traits may be indicative of adaptability and healthy socialization. Conversely, it has been suggested that individuals who blatantly distort their responses may also be more inclined to engage in other types of dishonest behaviors (see Lewicki, Poland, Minton, & Sheppard, 1997; Griffith, 2005; Rosse et al., 1998; Hakstian & Ng, 2005). This suggests that relying on social desirability scales is insufficient for accurately capturing these types of differences in response distortion, and especially for identifying applicants who engage in high magnitudes of faking behavior. As a result, using these scales may overlook important

individual differences in faking and obscure the potential impact of variability in faking across applicants with regard to criterion-related validity as well as job performance.

Given the weaknesses of the social desirability scales discussed above, it is clear that these scales may not be adequate for assessing applicant faking. Specifically, research has shown social desirability scales to correlate with personality measures, which may lead to increased rates of incorrectly labeling honest applicants as fakers. In addition, social desirability scales are unable to determine when an individual is providing a deliberately inaccurate response to a specific item. In response to these limitations of social desirability scales, another means of detecting faking that does not possess these weaknesses is needed. One method that has demonstrated promise, but little attention in the literature is the use of bogus item scales. The next section will discuss this approach in more detail.

Bogus Knowledge Scale

One recent approach to faking that has demonstrated promise in not only detecting the presence of faking but also in assessing individual differences in faking magnitude is the bogus knowledge scale (Anderson et al., 1984; Day & Carroll, 2008; Dwight & Donovan, 2003; Levashina, et al., 2009; Pannone, 1984). The bogus knowledge scale asks individuals to indicate their knowledge of or familiarity with fictitious concepts or items (e.g., "How familiar are you with the activities of the group *Citizens for Urban Crime Reduction*?"). Using a measure of fictional items offers a significant improvement over existing faking measures by providing a more accurate method of identifying individuals who are responding dishonestly and offering a way to detect variability in response distortion. When an applicant claims to be familiar with an

object or concept that does not exist, it is likely that he or she is responding dishonestly. While it is possible that an individual who is low in cognitive ability may be more likely to accidentally endorse a bogus item, this relationship has not been formally tested in the literature (Paulhus & Harms, 2004).

Using this type of scale in an applied setting is advantageous because it offers a more accurate measure of identifying individuals who are faking than other existing methods used to detect applicant faking. Specifically, social desirability scales are based on the assumption that high scores indicate a general tendency to distort responses, but they are unable to detect when a respondent has provided a specific, untrue response. The bogus knowledge scale, on the other hand, is able to provide insight about the frequency and extent to which applicants are distorting their responses. Using this type of scale in a selection context may help researchers and practitioners differentiate applicants who are responding honestly from those who are distorting their responses to get the job. Further, this scale is able to distinguish among applicants who are engaging in varying magnitudes of response distortion. Thus, while social desirability scales have been unable to differentiate exaggeration (or self-deception) from intentional response distortion (impression management), bogus items are able to capture intentional response distortion as well as assess the individual's magnitude of faking.

Although the bogus knowledge scale was developed several decades ago, it has been largely under-researched. The studies that have used this type of scale have incorporated the scale into job analysis surveys (Anderson et al., 1984), biodata inventories (Levashina, et al., 2009; Pannone, 1984) and job knowledge tests (Carroll, 2007; Dwight & Donovan, 2003). While the nature of the bogus knowledge item lends

15

itself to knowledge scales, researchers may consider modifying this type of item to fit the context of personality scales in the future. Incorporating bogus personality items may be worthwhile in light of research indicating that applicants who fake tend to do so across multiple measures (McFarland & Ryan, 2000).

Overall, studies utilizing bogus knowledge items have consistently shown the scale to be an effective means of assessing faking (Anderson et al., 1984; Pannone, 1984; Carroll, 2007; Dwight & Donovan, 2003; Levashina, et al., 2009). In a study by Anderson et al. (1984), applicants for 13 different positions were presented with a list of real and bogus job tasks. Of these tasks, 45% of the applicants indicated that direct experience with, or knowledge of, at least one of the nonexistent tasks. Another study by Pannone (1984) asked applicants for an electrician position to indicate whether they had ever used a bogus piece of electrical equipment. Consistent with Anderson et al.'s (1984) findings, Pannone found that 35% of applicants endorsed the bogus knowledge item. These studies provide initial support for the notion that many applicants are attempting to make a favorable impression by indicating they have job-related experiences that they could not possibly have had.

A study by Dwight and Donovan (2003) used two bogus task statements embedded in actual task statements to index faking, and found that the bogus items were useful for identifying fakers. They also found that people tended to score lower when they were warned not to fake. Paulhus, Harms, Bruce, and Lysy (2003) utilized a technique very similar to the bogus knowledge scale. Specifically, they created a measure called the Over-claiming Questionnaire (OCQ) and had participants indicate their knowledge of fictitious items in general categories (as opposed to items relating to a

specific job). Consistent with research conducted in an employment context, this study found that participants showed a substantially higher rate of over-claiming when trying to give a positive impression.

More recently, Carroll (2007) examined whether a bogus knowledge scale was a valid tool for identifying faking, and compared its effectiveness to that of an impression management scale. Employed individuals completed a bogus item scale as part of a laboratory study that involved a simulated employee selection situation. Participants' personalities were assessed through self- and peer-reports. Results demonstrated some support for the validity of the bogus knowledge scale in identifying response distortion, and indicated that the bogus knowledge scale was a better measure than an impression management scale. Specifically, a stronger correlation was found between admissions of faking and the bogus knowledge scale ($r = .38$) than the IM scale ($r = .18$).

Levashina, Morgeson, and Campion (2009) included three bogus knowledge items as part of a biodata inventory in a study investigating the relationship between job applicant mental abilities and faking. Results indicated that individuals who endorsed bogus items were also able to obtain higher scores on the biodata measure, and that applicants with high levels of mental abilities were less likely to engage in faking, or faked to a lesser extent in the selection context. The bogus knowledge items also effectively identified fakers and demonstrated that there are individual differences in faking. Specifically, the bogus items were endorsed differently across applicants, with some individuals endorsing one item and others endorsing multiple items. These findings suggest that the bogus knowledge scale is able to detect individual differences in faking by identifying faking on an item-by-item basis.

Overall, the research to date has shown the bogus knowledge scale to be an effective tool in identifying individuals who are faking (bogus item endorsement) as well as the extent to which these individuals are distorting their responses (the frequency and severity of bogus item endorsement). The research on the bogus knowledge scale suggests it has a number of strengths and advantages over social desirability scales. First, because bogus knowledge items reference fictitious constructs, they are unlikely to correlate with personality traits. While honest people can legitimately endorse some "unlikely virtues" items, they are not able to legitimately endorse bogus items (e.g., applicants cannot watch a television show that does not exist). Because bogus items are more verifiable, this may make applicants more reluctant to endorse them. If an applicant claims to have knowledge or experience in a particular area that seems job relevant, it would seem more plausible that they would have to rely on that knowledge or skill in a subsequent selection phase or on the job. From this frame of mind, applicants may be less willing to claim having skills or knowledges that they may have to demonstrate in the future. In fact, bogus items have been found to be more difficult for honest people to accidentally endorse (Kim, 2006). This finding is not surprising given the conceptual differences between bogus and social desirability items.

It is also possible that applicants may be more inclined to endorse social desirability items because they view them as less verifiable. Because social desirability items tend to be common behaviors, applicants may believe that responding dishonestly on these items will never be verified. From an applicant's perspective, it seems far-fetched that an organization would follow-up on specific responses to items regarding

how frequently they litter or curse. Further, applicants may view these behaviors as

irrelevant to job performance and not see the harm in distorting their responses to them.

Perhaps one of the biggest advantages of the bogus knowledge scale is that it has

the potential to detect intentional faking (Ramsay, Kim, Oswald, Schmitt, & Gillespie,

2008). Because bogus items capture respondents claiming to have impossible

experiences, they are able to measure conscious lying rather than other types of faking

(e.g., omission, exaggeration). Although it is possible that individuals may mistakenly

endorse one or two bogus knowledge items, individuals who endorse bogus items

frequently or claim to be very familiar with fictitious item content, are most likely

responding dishonestly. Additionally, because the bogus knowledge scale is able to

detect faking at an item-level, it is possible to identify the frequency and severity with

which an individual distorts responses. Using this item-level approach allows researchers

to calculate how many bogus items are endorsed and the level of endorsement (e.g., an

individual who states they watch a (fictitious) television program "one time a week" vs.

"4 or 5 times a week"), so that variance in faking can be assessed.

In spite of these advantages, there are a few limitations associated with the bogus

knowledge scale. First, research suggests that it would be relatively easy to coach people

to avoid these scales (Kuncel & Borneman, 2007). Second, in order for a bogus

knowledge item to be effective, respondents must realize that they are unfamiliar with the

construct, and then intentionally endorse the item in an attempt to appear more favorable

(Kim, 2006). Without both of these steps being fulfilled, an applicant's response may be

inaccurate, but he or she is not necessarily engaging in intentional faking. Kim (2006)

conducted a verbal cognitive interview with participants who had completed a bogus

knowledge scale, and found that several individuals indicated that they were somewhat confused by the item content. As a result, these individuals were confused when giving their responses, which may threaten the bogus item's ability to accurately identify intentionally inflated responses. Therefore, it is reasonable to expect that while bogus items will still be imperfect measures of faking, they may still be more valid than social desirability scales.

Despite these limitations, the bogus knowledge scale has several advantages that warrant future use to better assess applicant faking. Specifically, the bogus knowledge scale's ability to identify individuals who are intentionally responding dishonestly as well as detect the magnitude of response distortion offers an improved measure of applicant faking over social desirability scales. No other measure of faking has been able to provide such an accurate measure of frequency and severity in faking behavior, which makes using the bogus knowledge scale an improvement over existing measures.

Effects of Applicant Faking

As noted above, many questions still remain regarding the extent to which applicant faking impacts the validity and selection decisions of non-cognitive measures. Specifically, there is still no clear explanation as to why some studies have shown adverse effects (Douglas et al., 1996; Dunnette, et al., 1962; Griffith, 1998; Holden & Jackson, 1981; Mueller-Hanson, et al., 2003; Pannone, 1984; Rosse et al., 1998), while others have not (Abrahams, et al., 1971; Hough et al., 1990; Hough, 1998; McCrae & Costa, 1983; Ones, et al., 1996). One possible explanation, which is examined in this dissertation, is that previous research has not taken individual differences in applicant

faking into account. The issue of individual differences in faking and its potential impact on validity and selection decisions is explored in more detail in the next section.

Variance in Applicant Faking

To date, variability in faking has not been widely examined in the literature. The few studies that have examined individual differences in faking have suggested that people do fake to varying extents, and that there is variance even among fakers (Griffith, Chmielowski, Snell, Frei, & McDaniel, 2000; McFarland & Ryan, 2000; Rosse et al., 1998). Therefore, the current study examines differences in magnitude and variability in faking behaviors using the bogus item scale, which offers the ability to differentiate varying magnitudes of response distortion among individuals. Additionally, in order to examine the impact of faking on non-cognitive measures, it is important to verify that applicants engaging in extreme faking behavior on the bogus knowledge scale are also able to inflate their responses and obtain high scores on other selection measures.

Because the bogus knowledge items are not embedded within the biodata inventory in the current study, it is important to first investigate whether individuals who highly endorse the bogus items also have the high biodata scores. Examining this relationship will establish a link between faking on the bogus knowledge scale and response distortion on the biodata inventory. A study by McFarland and Ryan (2000) examined applicant faking across multiple measures, including several personality scales, an integrity test, and a biodata inventory. Results indicated that applicants who were able to inflate their responses on one measure were able to do so consistently across measures. In line with this research, it is expected that individuals in the current study who inflate their responses on the bogus knowledge scale will also inflate their responses on the

biodata inventory. These individuals should also be among the applicants with the

highest overall biodata scores. Thus:

Hypothesis 1: There will be a positive relationship between bogus item

endorsement and biodata score, such that applicants who engage in higher degrees

of faking behavior (as measured by bogus item endorsement) will have higher

overall biodata scores than those who engage in lower degrees of faking behavior.

Assessing faking variability across applicants is important for several reasons.

Given that a predictor measures an applicant on a trait that is expected to correlate with

job performance, the predictor score is assumed to be indicative of applicants who will

likely be successful on the job. If applicants fake to the same extent (i.e., there is no

variability in faking across individuals), the applicants with the highest predictor scores

would be expected to have the highest performance on the job. However, if applicant

faking does inhibit a test's ability to identify the most qualified applicants in a given pool

(i.e., applicants inflate their scores to varying extents and disrupt the true rank-order), the

criterion-related validity of the measure is undermined. Therefore, a closer look at the

potential impact of individual differences in applicant faking on the validity and quality

of selection decisions is warranted.

Applicant Faking and Predictive Validity

As noted above, the purpose of a selection test is to measure a certain ability or

trait in applicants that is expected to be positively related to job performance. If an

individual intentionally distorts his or her responses on a selection assessment and is

hired for the job based on these inaccurate responses, this should have implications for his or her task performance on the job, thereby affecting the predictive validity of the measure. In other words, if an assumption about an applicant's performance is based on inaccurate information (i.e., a faked test score), there should be a disparity between that individual's expected and actual performance. Additionally, the greater the disparity between the individual's reported and actual trait level on the predictor variable, the greater the expected disparity between the individual's assumed and actual performance. Because this disparity should be the largest among applicants who engage in high magnitudes of faking, the criterion-related validity for these individuals should be lower than those with low faking magnitudes (i.e., those who respond honestly).

This notion can also be explained from a measurement perspective. Komar et al. (2008) use Classical Test Theory (Crocker & Algina, 1986) as a framework to explain how applicant faking can adversely impact the criterion-related validity of non-cognitive. According to Classical Test Theory, test validity depends upon the extent to which tests scores contain true score variance relative to error variance (Crocker & Algina, 1986). The more a test score contains error or irrelevant constructs, the less it is able to predict the criterion of interest. Given that faking has been considered an artificial inflation of a test score, the amount of construct-irrelevant variance within a given test score should increase as faking magnitude increases. To the extent that faking is error variance (Smith & Ellingson, 2002) or at least unrelated to performance (as suggested by Ones et al., 1996), test validity should be increasingly undermined as faking magnitude increases (Rosse et al., 1998).

It has also been suggested that applicant faking can attenuate criterion-related validity by disrupting the true score rank-order of applicants (Rosse et al., 1998; Komar et al., 2008). Further, assuming there is variability in faking across applicants, the true rank-order of applicants should become increasingly disrupted as the amount of faking variability increases. In other words, the amount of true score variance that is related to performance and ranks applicants accordingly would become increasingly obscured as faking variability increases, deteriorating the predictive validity of the overall measure.

While most previous research has utilized a correlation coefficient to assess criterion-related validity, some researchers have suggested that examining the slope between a predictor and criterion is also necessary (Drasgow & Kang, 1984; Mueller-Hanson, et al., 2003). This has been recommended in light of criticisms that correlation coefficients can be insensitive to changes in rank-order in the extreme ends of trait distributions, which is where faking is expected to have the strongest effect (Drasgow & Kang, 1984; Levin & Zickar, 2002; Rosse et al., 1998; Zickar, Rosse, & Levin, 1996). In other words, while the observed criterion-related validity for an entire applicant pool may not change, there is reason to expect that the validity for applicants among the top end of the predictor distribution would still be deteriorated (based on the assumption that faking magnitude and variability would be highest among them) (Douglas et al., 1996; Levin, 1995; Zickar et al., 1996). In fact, research utilizing computer simulations have suggested that criterion-related validity is much more affected among groups with high magnitudes of faking (Converse et al., 2009; Komar et al., 2008; Rosse et al., 1998). Consistent with this research, a study by Mueller-Hanson, Heggestad, and Thornton (2003) found greater prediction error among students with the highest personality scores

in an incentive group, relative to those at the bottom portion of the personality score distribution. Specifically, this study found that among those in the incentive group, the criterion-related validity for individuals with predictor scores in the lower third of the distribution was significantly higher than the validity for those who scored in the upper third of the predictor distribution (.45 and .07, $p < .05$, respectively). Further, the criterion-related validities for those in the upper and lower thirds of the predictor distribution for the control group were not significantly different (.20 and .26, $p < .05$, respectively).

This research suggests that faking may differentially impact the criterion-related validity at different points on the predictor distribution. If this is the case, this is particularly troubling for researchers using only the correlation coefficient to assess the impact faking has on the predictive validity of a measure. Specifically, if faking is most likely to impact the rank-order at the top end of the predictor distribution and is not reflected by a change in the overall correlation coefficient, this has the potential to directly impact the quality of hiring decisions.

The research above suggests that applicant faking is most likely to affect the predictive validity of non-cognitive measures when 1) applicants engage in high magnitudes of response distortion, or 2) there is a high level of variability in faking behavior across applicants, resulting in changes to the rank-order of applicants. Therefore, the current study will examine the impact of both applicant faking magnitude and variability.

First, the impact of faking magnitude will be examined by comparing the criterion-related validity of applicants with higher faking magnitudes to those with lower

faking magnitudes. Based on the assumption that faking magnitude decreases the amount of true score variance of biodata scores, it is expected that individuals who engage in high faking magnitudes will have overall biodata scores that also contain the most error variance. This should result in a lower criterion-related validity for these applicants. Accordingly, it is expected that applicants who do not endorse bogus items (i.e., have a lower magnitude of faking) will have overall biodata scores with higher amounts of true score variance, resulting in higher criterion-related validities for predicting task performance as measured in an AC.

Figure 1 illustrates the expected relationships between biodata score and expected AC performance for applicants with low and high faking magnitudes (referred to as faking scores in the figures), respectively. Specifically, the first illustration depicts a scatter plot of data points that are expected to contain more true score variance and less faking variance. As a result, the predictive validity is higher among applicants with low faking scores. The second illustration displays data points for applicants with high faking magnitudes. These biodata scores are assumed to contain relatively more faking variance than true score variance, which attenuates the criterion-related validity among these applicants. (Note that Figure 1 is illustrative data and not actual data from a study.)

Next, this study will assess the impact of faking variability by examining the criterion-related validity at multiple points along the predictor distribution. Specifically, the criterion-related validity among applicants in the upper and lower ends of the biodata score distribution will be examined. As previously noted, research has suggested that faking variability will be highest among applicants at the top end of the predictor distribution (Drasgow & Kang, 1984; Levin & Zickar, 2002; Rosse et al., 1998; Zickar et

al., 1996) because these individuals have high biodata scores either because they are either truly qualified or are not truly qualified but have inflated their responses (Mueller-Hanson et al., 2003). Because there should be more variability in response distortion at the top end of the distribution list, the true rank-order of applicants is expected to be more disrupted, resulting in reduced criterion-related validity. Applicants with low biodata scores, on the other hand, are expected to be either truly less qualified than others in the applicant pool, or are unable or unwilling to inflate their responses. In both cases, these applicants' scores are expected to have minimal amounts of response distortion (or low faking variability), which should result in higher criterion-related validity.

Figure 2 presents an illustration of how faking variability is expected to impact the criterion-related validity at the top and bottom ends of the biodata score distribution. In looking at the lower end of the distribution line, the relationship between expected AC task performance and biodata score is stronger because these applicants are assumed to have low magnitudes of faking, which results in lower faking variability across applicants. However, the applicants at the top end of the biodata score distribution consist of those with high and low faking magnitudes. In other words, these applicants are either truly qualified and responded honestly or are not truly qualified and have inflated their scores. As a result, the criterion-related validity among these applicants is expected to be reduced by the high variability across applicants.

Hypothesis 2a: Faking will impact the criterion-related validity of assessment center task performance, whereby the predictive validities will be higher for honest respondents than extreme fakers.

Hypothesis 2b: The predictor-criterion relationship will be stronger among applicants in the bottom portion of the biodata score distribution than applicants in the top portion.

The research discussed above poses a potential concern regarding the quality of selection decisions using noncognitive measures. If there are no detectable effects on the overall validity coefficient, researchers and practitioners relying on this estimate may minimize or dismiss the potential consequences associated with applicant faking. As a result, response distortion may have a dramatic effect on who is hired in that the applicants who are likely to be hired are likely to have faked to some degree in order to get the job. However, hiring individuals who have inflated their responses is not necessarily problematic, unless these individuals also are poorer performers than those who provide an honest representation of themselves. Therefore, the next issue to be examined is whether hiring fakers has implications for subsequent task performance, as measured in an AC.

Applicant Faking and Performance

Having valid selection tools is important for ensuring organizations are selecting the most qualified applicants for the job. If applicants are misrepresenting themselves on a selection test, the risk of hiring people who are not qualified for the job increases. Therefore, the real concern with applicant faking as it pertains to job performance seems to be whether applicants who severely distort their responses are able to perform the job tasks successfully.

There has been much speculation within the literature regarding the relationship between faking and job performance. Some researchers have suggested that faking has little to no impact on performance (Hogan, Barrett, & Hogan, 2007; Hogan, 1991; Hough et al., 1990; Viswesvaran et al., 1996; Viswesvaran, Ones, & Hough, 2001). Hogan (2006) has proposed that people treat social interactions similarly to selection contexts: by trying to present a positive image. In this view, it is nearly impossible to distinguish faking from socialized behavior, which is to say that faking would not necessarily impact future job performance. A meta-analysis by Ones et al. (1996) examined the relationship between socially desirable responding and job performance. Results indicated that social desirability scores were uncorrelated with measures of task performance ($\rho = .00$), job performance ($\rho = .01$), and counterproductive behavior ($\rho = -.03$), although they were modestly related to ($\rho = .19$) measures of training performance. However, multiple criterion measures (e.g., different definitions of job performance) as well as various conceptualizations of social desirability (e.g., studies focusing on impression management versus self-deception) were included in the meta-analysis, which limits the reliability of these results. Additionally, Ones et al. did not consider type of job as a moderator of the social desirability-performance relationships. Given these limitations, the results should be interpreted with caution.

Some researchers have suggested that the ability to distort scores in a socially desirable direction indicates a functional awareness of social norms, which might in fact be related to higher levels of performance (e.g., Hogan et al., 1996; Pauls & Crost, 2005). While having the ability to adapt to social expectations in some instances may be beneficial, several other researchers have suggested that applicants who misrepresent

themselves in selection contexts are less qualified than honest respondents (McFarland &

Ryan, 2000, 2001; Mueller-Hanson, et al., 2006; Rosse, Levin, & Nowicki, 1999; Zickar,

Rosse, Levin, & Hulin, 1996). For instance, Rosse et al. suggested that for certain

positions, the tendency to give incorrect information may be highly dysfunctional, and

may be counterproductive for developing effective long-term relationships. Given these

discrepant findings, more empirical evidence is needed to better understand how faking

in a selection context translates to job performance.

There are several reasons to assume that faking itself may be related to poor task

performance. One possible explanation is that applicants who present a false impression

about themselves to get a job do so because they are not qualified to perform the job

tasks. In other words, there is a discrepancy between the individual's true abilities and

those that are needed to perform the job successfully. Further, as the magnitude of

response distortion increases, so will the difference in expected and actual performance.

Given that the purpose of selection tests is to identify which applicants have the

characteristic(s) and/or knowledge(s) that are important to the job, applicants who

severely misrepresent themselves should not be able to perform the job tasks as well as

applicants with those qualities. Therefore, it is possible that the impact of faking on job

performance may depend on the magnitude of the distortion, such that individuals who

engage in greater magnitudes of faking behavior will have the lowest levels of task

performance.

Another possible explanation as to why fakers are expected to be poor performers

is because they possess personality characteristics that are associated with both faking

and poor performance. In other words, the personality characteristics that lead a person

to fake on a selection test may also cause that individual to engage in work behaviors that are associated with poor task performance. Research examining the various factors that contribute to applicant faking have indicated that dispositional factors are among the variables related to the ability and motivation to fake (McFarland & Ryan, 2000, 2001; Mueller-Hanson et al., 2006; Snell et al., 1999). Further, this research suggests that the extent to which an individual fakes is partially determined by personality characteristics (McFarland & Ryan, 2000; Mueller-Hanson et al., 2006). Some of the dispositional characteristics that have been examined with regard to applicant faking are conscientiousness, neuroticism, and cognitive ability (Huws, Reddy & Talcott, 2009; McFarland & Ryan, 2000, 2001; Mueller-Hanson et al., 2006; Snell et al., 1999), which are characteristics that have shown consistent relationships with overall job performance (Barrick & Mount, 1991; Barrick et al., 2001; Hough, 1992; Gatewood & Field, 2001; Ones & Viswesvaran, 2001; Ones, Viswesvaran & Schmidt, 1993; Salgado, 1997; Schmidt, Shaffer & Oh, 2008; Schmidt & Hunter, 1998; Tett et al., 1991).

Conscientiousness has been associated with integrity (Ones et al., 1993), suggesting that individuals who are high in this trait also have a tendency to be honest (Costa & McCrae, 1989). McFarland and Ryan (2000) also found that conscientiousness was negatively correlated with faking behavior. With regard to job performance, conscientiousness has consistently shown a positive relationship with performance across jobs and performance criteria (Barrick & Mount, 1991; Barrick et al., 2001; Salgado, 1997; Tett et al., 1991).

Research has shown that individuals with high levels of neuroticism tend to be very concerned with social appropriateness and how others view them (Costa & McCrae,

31

1989), which may lead these individuals to try to impression manage by engaging in response distortion. McFarland and Ryan (2000) also found that neuroticism was positively correlated with faking behavior. Neuroticism has also been shown to have a negative relationship with performance across jobs and performance criteria (Barrick & Mount, 1991; Barrick et al., 2001; Salgado, 1997; Tett et al., 1991). Neurotic individuals tend to be anxious, angry, worried, and insecure (Borgatta, 1964; Conley, 1985; Hakel, 1974; John, 1989; Lorr & Manning, 1978; McCrae & Costa, 1985; Noller et al., 1987; Norman, 1963; Smith, 1967), which likely hinders them from performing job tasks successfully.

Research investigating whether cognitive ability influences faking on personality measures has offered inconsistent findings. While some researchers have found that the ability to fake was not correlated with intelligence (Anderson et al., 1984; Burkhart et al., 1978; Garry, 1953; Law, Mobley & Wong, 2002; Ones et al., 1996), other studies have reported significant relationships between intelligence and the response distortion of non-cognitive measures (Alliger et al., 1996; Grubb & McDaniel, 2007; Nguyen, 2001; Pauls & Crost, 2005). Several non-experimental studies have suggested that more intelligent people are less likely to fake or they fake to less extent (Austin et al., 2002; Egan, 1989; Eysenck, 1971; Levashina, et al., 2009; Nguyen, 2001).

Levashina, Morgeson, and Campion (2009) proposed that one reason for this negative correlation is that individuals with high abilities are less willing to cheat because they expect to do well even without cheating. In other words, individuals who are confident in their knowledge and abilities are probably less likely to intentionally misrepresent themselves. On the other hand, individuals who are less confident in their

abilities may be more willing to distort their responses in an attempt to make a better impression of themselves. Cognitive ability has also been shown to be related to a number of important work-related outcomes, such as job performance (Hunter, 1986; Hunter & Hunter, 1984; Schmidt & Hunter, 1998), job knowledge and success at training (Schmidt & Hunter, 1992), and counterproductive work behaviors (Dilchert, Ones, Davis, & Rostow, 2007). Overall, cognitive ability has consistently demonstrated a positive relationship with general performance (Schmidt, et al., 2008; Gatewood & Field, 2001; Schmidt & Hunter, 1984).

In sum, the research discussed above suggests that the individuals who are more likely to engage in response distortion are those who have low conscientiousness, high neuroticism, and low cognitive ability (Huws et al., 2009; McFarland & Ryan, 2000, 2001; Mueller-Hanson et al., 2006; Schab, 1991; Snell et al., 1999). Further, this research has shown these traits to be independently related faking behaviors. These characteristics have also demonstrated a consistent negative relationship with general job performance (Barrick & Mount, 1991; Barrick et al., 2001; Gatewood & Field, 2001; Hough, 1992; Ones & Viswesvaran, 2001; Ones et al., 1993; Salgado, 1997; Schmidt et al., 2008; Schmidt & Hunter, 1984; Tett et al., 1991).

It is worth noting that while faking is likely to differentially impact various aspects of job performance (i.e., contextual and counterproductive performance), the current study focuses only on task performance. Specifically, the performance criterion used in the current research was an AC, which consisted of three exercises (individual exercise, group exercise, and a written demonstration). ACs are a commonly used tool that contain a variety of performance measures that are used for assessment and

development purposes. ACs are typically performance-oriented in that applicants complete a variety of exercises that are to demonstrate a variety of skills that are believed to be related to job performance. Research has demonstrated relatively favorable validity evidence for AC dimensions, with criterion-related validities in the .36 range for AC rating and job performance (Arthur, Day, McNelly, & Edens, 2003; Bowler & Woehr, 2006; Gaugler, Rosenthal, Thornton, & Bentson, 1987; Meriac et al., 2008). Therefore, there is empirical support for using AC performance score in the current study.

Given that in the current study, individuals participated in the AC as the final stage of the selection process, it is possible that some individuals will engage in impression management (IM) behaviors. Only one study has examined IM in ACs, which found that individuals used more IM tactics in interactive exercises (e.g., role-play), and that IM tactics impacted the assessors' evaluations (McFarland, Yun, Harold, Viera, & Moore, 2005). While it is possible that applicants in the current study will also engage in IM behaviors, its impact on raters' evaluations is less concerning given that the assessors were trained by industrial-organizational psychologists in assessing performance.

As stated above, there are two possible reasons to expect a negative relationship between extreme faking and task performance. First, applicants who have to severely distort their responses in an attempt to meet the qualifications for the job may not be capable of performing the job tasks as well as honest respondents with those necessary qualities. Second, the dispositional characteristics that may lead applicants to engage in extreme response distortion might also be related to poor job performance. Given that most individuals who engage in extreme response distortion are more likely to end up at

the top end of the biodata score distribution, it is expected that there will be a greater disparity between assumed and actual task performance among these individuals. Therefore, the issue of applicant faking in a selection context is most concerning when it comes to hiring decisions. Assuming that organizations will use a top-down selection process to hire applicants, applicants at the top end of the predictor distribution are the most likely to be selected first. If there are performance differences among applicants who engage in high magnitudes of response distortion and those who do not, this is likely to have the most drastic effect with regard to performance among applicants at the top end of the predictor distribution because there should be the largest discrepancy between expected and actual performance among these individuals.

Therefore, a final goal of this dissertation is to examine whether there are performance differences, as measured in the AC, among individuals who engage in high and low magnitudes of faking at the top end of the biodata score distribution. Specifically, within the top portion of the biodata score distribution (those who would be most likely to be hired first in a top-down selection process), AC task performance scores of individuals who engaged in higher degrees of faking (as measured by bogus item endorsement) will be compared to individuals with low faking magnitudes. It is expected that among the applicants in the top of the score distribution, individuals who endorsed more bogus items will have lower AC task performance scores than those who engaged in low faking magnitudes. Thus:

> *Hypothesis 3.* Within the top third portion of the biodata score distribution, the
> mean AC task performance among applicants in the high faking magnitude group

will be lower than the AC task performance of applicants in the low faking

magnitude (honest) group.

Summary

This research contributes to the applicant faking literature in multiple respects.

First, a novel measure of applicant faking is used to better understand the prevalence and

variability of faking in an applicant sample. Second, this study provides a potential

explanation as to why existing research has found conflicting findings regarding the

extent to which applicant faking affects the criterion-related validity of non-cognitive

measures. Finally, this study examines performance data to determine whether existing

beliefs about fakers being poor performers is valid.

Chapter 3: Method

Participants

Participants were 970 applicants for the 2006 *Federal Presidential Management Fellows* (PMF) program, which is managed by the United States Office of Personnel Management (OPM). A total of 1,090 applicants were given a voluntary research test battery containing the measures of interest (i.e., the biodata inventory and bogus item scale) and 970 applicants (89.0%) completed it. Because the research battery was administered during the formal application process into the PMF program, demographic data was not collected and cannot be reported.

This sample was appropriate for the current study because it was a real-world applicant pool that consisted of individuals who were highly motivated, and therefore, more inclined to engage in faking behavior (Schmit & Ryan, 1993). These individuals were highly qualified (e.g., all had graduate degrees) and motivated to perform well on the assessments. Furthermore, the PMF program is a highly competitive and widely renowned program within the Federal government. Given the incentive associated with becoming a PMF, it is likely that the applicants were highly motivated to perform well, which may have led some individuals to engage in the faking behaviors of interest to this study.

The PMF program was established by Executive Order in 1977 to attract to the Federal service individuals from a variety of academic disciplines who have an interest in, and who are committed to, a career in the leadership and management of public policies and programs. All program participants were U.S. citizens who either had or

were in the process receiving a graduate degree. The selection process for the PMF program occurred in four stages, which are described below in more detail.

In the first stage of the selection process, universities and institutions nominated one individual for consideration as a PMF. Although it is up to the institution itself to determine who is nominated, OPM required that each institution ensure that the individual had a high level of interest in government service, had a breadth and quality of accomplishments, and a capacity for leadership.

In the second stage, all individuals who were nominated for the PMF program were required to respond to three items on an Accomplishment Record (AR). The AR items were designed to assess the individual's competency in three dimensions: Interpersonal Skills, Problem Solving, and Resilience. Applicants were instructed to write a brief narrative describing a personal achievement that best represented their competence in each competency area. Two trained OPM assessors then rated the ARs. Applicants who obtained a certain score or higher on the ARs advanced to the Assessment Center (AC, stage 3).

All applicants who moved on to stage three, the AC, began on equal ground. Thus, finalists for the PMF program were selected only based on their performance in the AC. The AC consisted of three exercises: an individual presentation, a group discussion, and a written demonstration. The five competencies assessed in the AC exercises were: Flexibility, Interpersonal Skills, Oral Communication, Problem Solving, and Written Communication. The AC is discussed in more detail later in the Measures section.

Applicants who made it past stage three were designated as a "PMF Finalist" and moved on to the fourth stage, Agency Placement. In this stage of the selection process,

PMF Finalists participated in a job fair and other hiring activities with Federal agencies. However, PFM Finalists were not guaranteed placement in a Federal agency.

Procedure

In addition to the AC in stage three of the PMF selection process, applicants were asked to fill out an exploratory test battery containing a variety of assessments, including a biodata inventory and a job knowledge test. Completing the test battery was voluntary. Applicants were told that whether or not they completed the test battery would in no way affect their advancement in the PMF process, and that the purpose of the battery was to gather data to improve the PMF process. Packets were given to applicants at the beginning of the AC and were instructed to complete the measures over the course of the AC.

Since completion of the assessment battery was voluntary, it is possible that completion of the battery during the AC may have affected these applicants' performance on the AC itself. A manipulation check was conducted to ensure that there were no differences in AC score between those applicants who completed the battery and those who did not. Specifically, an independent samples t-test was used to compare the mean AC scores of the two groups. Results indicated that there was no significant difference between those applicants who completed the battery and those who did not, $t(1088) = .35, p = .72$. This finding suggests that completing the battery did not impact applicants' performance in the AC.

Measures

Bogus Item Scale. This dissertation used a bogus item scale (see Appendix A; bogus items are bolded) that was developed by OPM Ph.D. psychologists to examine

faking among applicants. Bogus and authentic items were written covering a wide range of topic areas that were deemed relevant to PMF applicants' potential jobs. Specifically, the item content included world, national, and government knowledge. To verify that the bogus items were, in fact, nonexistent aspects of the job, subject matter experts independently reviewed the bogus items. Additionally, internet searches were conducted to ensure that the content of the bogus items did not have other potential meanings and could not be unambiguously interpreted as bogus (e.g., *Pan Arab Alliance for Peace in the Middle East*).

The bogus item scale consisted of 15 items, with eight of the items referencing real concepts and seven containing bogus content. The scale was designed to resemble a job knowledge questionnaire. Applicants were asked to rate their knowledge of an item using either a *familiarity scale* [(A) = Very Familiar, (B) = Somewhat Familiar, (C) = Familiar, (D) = Vaguely Familiar, (E) = I have never heard of (topic)] or a *frequency scale* [(A) = 1, (B) = 2 to 3, (C) = 4 to 5, (D) = less than once a week, (E) = I have never heard of (topic)], depending on the topic described in the item. The bogus item scale was based on the rationale that individuals who were faking would indicate that they were familiar with topics that did not exist, whereas individuals who were responding honestly would not indicate that they were familiar with the fictitious item content. Therefore, items were recoded so that higher scores reflected higher levels of familiarity or frequency with the item content (i.e., 0 = Not Familiar, 1 = Vaguely Familiar, 2 = Somewhat Familiar, 3 = Familiar, and 4 = Very Familiar). A total score was then calculated by summing all responses to the seven bogus items, such that higher scores

indicated that the applicant endorsed bogus items at a higher frequency or severity (Carroll, 2007).

This measure also included items that contained real content. These items were scored similarly to the bogus knowledge items, with higher scores indicating higher levels of familiarity. Although the measure correlated with the bogus item scale ($r = .38$, $p < .01$), it did not correlate with AC score so it was not included in the research hypotheses.

Because bogus items are intended to capture intentional response distortion, there should be consistency in the way applicants respond. The Cronbach's alpha for the Bogus Item Scale was .66. The Total Bogus Score measure was used as a basis to categorize applicants into one of three response groups: Honest Respondents, Moderate Fakers and Extreme Fakers. Because the bogus item scale is relatively uncommon in the literature, there is limited guidance in terms of how to categorize varying degrees of faking. Of the research that has used bogus items, two studies have sought to categorize individuals into more than two types of respondents (honest versus fakers). Levashina et al. (2009) used three bogus items and categorized respondents into four groups based on the number of items the applicant endorsed (zero, one, two, or three). Alternatively, Dwight and Donovan (2003) calculated a total bogus score based on responses to two bogus items that used a 4-point frequency scale (0-3) and applied two cut points. Individuals who obtained a score of four or higher (by endorsing at least two bogus items at a severity of at least two) were identified as extreme fakers, and individuals with a bogus score of at least two were classified as moderate fakers. Given that this dissertation used more bogus items and a larger scale than these studies, which resulted in

a larger range of total bogus item scores, the current study used a one standard deviation split to divide individuals into three categories. Specifically, applicants who scored more than one standard deviation above the mean Total Bogus Score were categorized into the "Extreme Faking Group," and individuals with scores more than one standard deviation below the mean score were categorized into the "Honest Respondent Group." All other individuals were categorized into the "Moderate Faking Group."

Creating groups using a standard deviation split was used for several reasons. First, a one standard deviation split has been recommend to create low and high scoring groups (Aiken & West, 1991). Second, it is reasonable to expect that individuals who frequently endorsed bogus items did so at a higher severity, based on research suggesting that individuals vary in the extent they are willing to engage in faking behavior (Lueke, Snell, Illingworth, & Paidas, 2001; McFarland & Ryan, 2000; Muller-Hanson et al., 2006). In the current study, a significant correlation ($r = .49$, $p < .01$) was found between frequency and severity of bogus item endorsement. This correlation supports the notion that applicants with a high Total Bogus Score tended to engage in more response distortion. In other words, applicants who frequently endorsed bogus items were more likely to do so at a higher severity, and would be in the top portion of the faking score distribution. Finally, in order for an applicant to receive a Total Bogus Score high enough to be placed in the "Extreme Faking Group," he or she would have indicated familiarity (a severity endorsement of two out of four) with the item content of at least two bogus items. This type of response pattern is more indicative of intentional response inflation rather than confusion or accidental endorsement (Paulhus & Harms, 2004; Ramsay et al., 2008).

Biodata inventory. The biodata inventory asked applicants about their educational background, work, and other life experiences. To ensure content validity of the measure, OPM psychologists reviewed and linked the items to competencies identified as critical across leadership positions. Specifically, the items targeted a variety of individual difference variables (e.g., personality traits, cognitive ability, and work-oriented competencies). The inventory consists of 31 multiple choice and likert-type format questions. Ten of the biodata items were categorical, and the content of these items did not have a clear or theoretical relationship with the PMF program. Additionally, the internal consistency of these ten items did not reach an acceptable level of internal consistency and did not demonstrate a significant relationship with the dependent variable, AC score. For these reasons, the ten categorical items were removed from the biodata inventory. Removal of these items enhanced the reliability of the remaining items and strengthened the relationship with AC score. The total biodata score was based on 21 items that utilized a likert-scale response scale; these items were scored using an empirical-rational scoring key. The bogus items were not included in the biodata total score. The Cronbach's alpha for the biodata inventory was .66, which falls shy of the recommended alpha of .70 (Nunnally, 1967) and is lower than the observed alphas in similar research (Carroll, 2007; Levashina et al., 2009). One possible explanation for the low reliability in this study is the heterogeneity of the measure, which is inherent with biodata measures. Additionally, the current measure is much shorter than those used in similar research, which may have also contributed to a lower alpha.

Assessment Center (AC). As mentioned previously, the PMF AC consisted of three exercises, including an individual presentation, a group discussion, and written

demonstration. Five competencies were assessed within the three exercises: Flexibility, Interpersonal Skills, Oral Communication, Problem Solving, and Written Communication. Note that Flexibility, Oral Communication, and Problem Solving were assessed in both the individual presentation and the group discussion. Trained OPM assessors assessed each competency in the AC on a 1-5 scale.

For the individual presentation, candidates were provided a one-page summary of an issue and were instructed to prepare a five-minute presentation describing their position on the issue. Each applicant was given 25 minutes to prepare his or her presentation that was given in front of a panel of assessors. After the five-minute presentation, a panel member presented an additional problem, or an "intervention," to which the applicant had two minutes to prepare a response. The panel of assessors observed and scored the presentation and additional response to the "intervention." Assessors rated the candidates on Problem Solving, Oral Communication, and Flexibility. After the panel came to individual ratings independently, assessors shared their individual ratings. If assessor ratings were more than one point apart (e.g., 3, 4, and 5; 2, 2, and 4), assessors engaged in a consensus discussion until all assessors were within one point of each other (e.g., 3, 4 and 4; 2, 2, and 3). Applicants' final competency scores were the panel's majority consensus score (e.g., if the assessor panel scores were 2, 2 and 3 for a competency, the applicant's final score for that competency would be 2).

After the individual presentation, applicants were assembled for the group discussion exercise. Group exercises consisted of four to six applicants, depending on the number of applicants being assessed that day. In the group discussion exercise, applicants were presented with the same issue presented to them in the individual

presentation. Applicants were instructed to briefly state their position on the issue, discuss the issue, and decide on a recommendation regarding how the issue should be resolved. Once the group had decided on a recommendation, they presented it to an assessor panel. The amount of time given to a group to discuss the issue and come to a recommendation depended on the number of group participants. The panel of assessors observed and took notes on the behavior of the applicants as they presented. Applicants were rated on Flexibility, Interpersonal Skills, Oral Communication, and Problem Solving in this exercise. Assessors went through the same consensus process for the individual presentation exercise as for the group exercise (described above).

For the written exercise, applicants were given 45 minutes to complete an essay. The essay was structured to allow applicants to demonstrate their competence in written communication. A different set of assessors rated the written exercise to ensure that the ratings were independent of an applicant's performance in the other AC exercises. Assessors engaged in a consensus discussion when their ratings differed by at least one point (e.g., 2 and 3). Assessors discussed their individual ratings and came to consensus so that their ratings were in perfect agreement (e.g., 3 and 3).

Final AC scores were computed by summing the final competency scores from each of the three exercises of the AC process for each applicant. For competencies measured in more than one exercise, the scores for each exercise were averaged for that competency. For example, if an applicant received a '2' on Problem Solving for the

individual presentation and a rating of '4' on Problem Solving for the group discussion,

that applicant received a score of '3' for the Problem Solving competency[1].

[1]Intra-class correlations ICC(2, k) were conducted to assess the consistency and agreement of assessor scores. Results were statistically significant, indicating substantial to high agreement among assessors across competencies (0.75 to 0.94, $p < .001$).

Chapter 4: Results

Table 1 provides the descriptive statistics, correlations, and coefficient alpha

reliabilities of the variables used to test the hypotheses. Nine percent of respondents did

not endorse any bogus items; 14% of respondents endorsed at least one bogus item; 21%

of respondents endorsed 2 bogus items; 20% of respondents endorsed 3 bogus items;

16% of respondents endorsed 4 bogus items; 10% of respondents endorsed 5 bogus

items; 6% of respondents endorsed 6 bogus items; and 4% of respondents endorsed all 7

bogus items. Additionally, the bogus items were endorsed in different ways. Bogus item

1 was endorsed by 51% of respondents, bogus item 2 was endorsed by 74% of

respondents, bogus item 3 was endorsed by 45% of respondents, bogus item 4 was

endorsed by 20% of respondents, bogus item 5 was endorsed by 21% of respondents,

bogus item 6 was endorsed by 28% of respondents, and bogus item 7 was endorsed by

47% of respondents. The average severity of endorsement (on a scale of 0-4) was .82 for

bogus item 1, 1.20 for bogus item 2, 0.69 for bogus item 3, 0.27 for bogus item 4, 0.25

for bogus item 5, 0.34 for bogus item 6, and 0.75 for bogus item 7.

Hypothesis 1 stated that applicants in the Extreme Faking Group would obtain the

highest scores on the biodata measure, and that applicants in the Honest Respondent

Group would have the lowest scores on the biodata measure. To test this hypothesis, a

one-way ANOVA was conducted with Faking Group as the independent variable and

total biodata score as the dependent variable. The results of the one-way ANOVA

indicated that a significant response category effect $F(2, 967) = 8.70, p < .01$.

Additionally, the mean biodata score increased in the expected direction (with individuals

in the Extreme and Moderate Faking Groups having higher mean biodata scores than the

Honest Respondent Group). The results of the Tukey-Kramer post hoc analysis indicated

that the Extreme Faking Group mean ($M = 92.81$, $SD = 5.72$) was significantly different

($p < .05$) from the Honest Group ($M = 90.69$, $SD = 5.51$) and the Moderate Faking Group

($M = 90.62$, $SD = 5.65$). The scores for the Extreme Faking Group were nearly one third

of a standard deviation higher than those for the Honest Respondent and Moderate

Faking Groups ($d = -.38$ and $d = -.39$, respectively), which is a small to moderate effect

(Cohen, 1992). However, there was not a statistically significant difference in the mean

biodata scores for the Honest Respondent and the Moderate Faking Groups ($d = .01$).

Table 2 provides information on the means, standard deviations, and effect sizes of the

biodata inventory scores across the different faking groups. Based on these findings, H1

was partially supported[2].

Hypothesis 2a stated that Faking Group would impact the criterion-related

validity of AC performance, with higher predictive validities for Honest Respondents

than Extreme Fakers. Response group aside, there was a small but statistically significant

correlation between Biodata score and AC performance, $r = .12$, $p < .01$ (observed

correlation). Hypothesis 2a was tested using regression analysis, with biodata score

predicting AC performance score and Faking Group as a moderating categorical variable.

Regression lines for the prediction of AC performance score were compared across

Faking Groups through Potthoff analysis, which provides a simultaneous test of slope and

intercept values. The results found no significant differences between the slope or

intercepts across the response groups, indicating that the regression lines did not

---

[2] Power analyses were conducted for all hypothesis tests. The results indicated that the current sample size was more than adequate to detect all hypothesized relationships with a minimum power of .80 and a Type I error rate of .05.

significantly differ between the response groups. These results do not provide support for

Hypothesis 2a. Regression analyses for this hypothesis are presented in Table 3.

Although the regression lines for biodata predicting AC scores were not found to

be significantly different across response groups, it is still possible that magnitude of

bogus item endorsement can affect the prediction of AC scores. Therefore, a multiple

regression was conducted with total bogus score as a continuous variable. Results

indicated that although the interaction term between biodata score and bogus item

endorsement was not significant, biodata score, $b = .14$, $t(967) = 4.17$, $p < .01$, and extent

of bogus item endorsement, $b = -.08$, $t(967) = -2.54$, $p < .05$, were significant predictors

of AC score. This indicates that faking magnitude did contribute to the prediction of AC

performance scores (albeit to a minimal extent); however, this effect was lost when the

variable was categorized. This loss of an effect occurs because continuous variables

allow for the maximization of variance in a scale, which is reduced when the variable is

categorized. The regression analyses can be found in Table 4.

Hypothesis 2b stated that the predictor-criterion relationship would be

significantly stronger in the bottom portion of the biodata score distribution than in the

top portion. To evaluate Hypothesis 2b, respondents were divided into thirds based on

their biodata scores. Next, the validity coefficients of scores from applicants in the upper

and lower third of each portion were compared. Among applicants in the bottom third of

the biodata distribution, the correlation between biodata inventory and AC score was .10

($p > .05$); this correlation was .06 ($p > .05$) among applicants in the top third of the

biodata distribution. These correlations were not statistically different ($z = .50$, $p > .05$).

While these results do not provide support for Hypothesis 2b, the correlation size is in the

expected direction; therefore, it is still a possibility that faking may differentially impact the criterion-related validities at various points in the biodata distribution.

Hypothesis 3 stated that within the top third portion of the biodata score distribution, performance among applicants in the Extreme Faking Group would be lower than the mean AC performance scores of those in the Honest Respondent Group. To test this hypothesis, a one-way ANOVA with Faking Group as the independent variable and AC performance score as the dependent variable was conducted for applicants in the top third of the biodata distribution.

The results of the one-way ANOVA revealed that there was not a significant response category effect, indicating that the mean AC scores between the Honest Respondent Group ($M = 17.83$, $SD = 2.37$) and the Extreme Faking Group ($M = 17.97$, $SD = 2.80$) were not statistically different. The difference in mean AC scores between these two groups was minimal ($d = .05$); therefore, Hypothesis 3 was not supported[3]. Table 5 provides means, standard deviations, and effect sizes AC performance scores for the Honest Respondent Group and Extreme Faking Group.

---

[3]Analyses were also conducted to determine whether there were performance differences at the competency level. No significant differences in AC competency scores were found across response groups for any competency.

Chapter 5: Discussion

Major Findings

The primary goals of the current study were to 1) examine the extent to which individual differences in faking impacts the criterion-related validity of non-cognitive measures in predicting task performance in an Assessment Center, and 2) determine whether there are differences in AC task performance scores among individuals who distort their responses on a biodata inventory relative to those who respond honestly. To investigate these objectives, a bogus knowledge scale was used to assess individual differences in faking magnitude and variability in an applied setting. Although the overall results were not uniformly supportive of the notion that individual differences in faking impacts criterion-related validity, some of the findings were encouraging nonetheless. There was also little support for the notion that there are differences in AC task performance among individuals who engage in faking behaviors and those who do not. However, the current study contributes to the applicant faking literature in several respects.

First, this study offers a better understanding of applicant faking behaviors in a real-world context by demonstrating that applicants who endorse bogus items also inflate their responses on non-cognitive measures. While the finding that applicants can and do fake is not new news, the results for Hypothesis 1 indicate that those who engage in faking behaviors do so across measures. Specifically, applicants who endorsed bogus items more frequently and at higher magnitudes obtained the highest biodata scores. Although the difference in mean biodata scores was only significant for the Extreme Faking Group, the pattern of biodata scores is consistent with expectations and previous

literature indicating that individuals inflate responses across measures (Levashina et al., 2009; McFarland & Ryan, 2000); however, these findings are not definitive with respect to the direction of causality.

The results from Hypothesis 1, along with this previous research, suggest that to the extent that faking is irrelevant error, inflated non-cognitive predictor scores should be less meaningful. Therefore, Hypothesis 2a sought to examine the extent to which error variance in non-cognitive predictor scores impacts the criterion-related validity of non-cognitive measures in predicting task performance in an AC. Results for Hypothesis 2a indicated that faking magnitude did not impact the predictive validities across response groups.

Although the results for Hypothesis 2a were not supported, this study still contributes to the debate regarding whether faking impacts the predictive validity of self-report measures. Although the findings for H2b were not significant, the pattern of correlations did suggest that variability in faking may affect criterion-related validities at various points in the predictor distribution. Specifically, the correlation coefficient for applicants in the bottom third of the biodata distribution was slightly higher than the correlation coefficient for those in the top third. Although these results did not reach statistical significance, the pattern of correlations is consistent with previous research (Mueller-Hanson et al., 2003) and suggests that variability in faking may differentially affect criterion-related validities at various points in the predictor distribution – namely, at the top of the predictor distribution. This could have important implications for organizations utilizing a top-down selection process, which may be more likely to select an individual with an inflated biodata score over an applicant who responded honestly.

However, hiring fakers over honest respondents is not necessarily problematic unless there are performance differences associated with varying amounts of applicant faking.

Hypothesis 3 investigated whether there were differences in applicant AC task performance scores across faking groups in the top third of the biodata score distribution. Only applicant scores at the top end of the predictor distribution were examined because these are the individuals who would be selected first using top-down selection. Results indicated that there were no mean differences in AC task performance across response groups in the upper portion of the predictor score distribution. These findings suggest that to the extent that the AC excises reflect tasks performed on the job, hiring extreme fakers over honest respondents would have minimal impact in terms of task performance.

The lack of observed performance differences is contrary to expectations. It was hypothesized that individuals who engaged in faking behaviors were claiming to have higher levels of characteristics that were associated with successful performance. The lack of support for this hypothesis suggests researchers continue to examine the theoretical relationship between faking and performance. For instance, Hogan's (1991) socioanalytic theory of personality states that faking is an applicant's means of telling a potential employer that they want to be someone who is highly conscientious or highly ambitious. The theory stipulates that these individuals are motivated to act in alignment with these claimed traits. Applicants in the current study completed the biodata inventory during the AC. It is possible that they were more conscious about ensuring that their behaviors in the AC were consistent with their responses on the biodata inventory, which resulted in minimal differences in AC task performance across faking groups.

The current study also contributes to the research examining attributes of individuals who engage in response distortion. For instance, an interesting observation was the large number of individuals who frequently endorsed bogus items and at high magnitudes. Given that all of the applicants either had a graduate degree or were in the process of obtaining one, it is plausible that these individuals are relatively conscientiousness and have higher levels of cognitive ability, in general. Based on these characteristics, existing research would suggest that these individuals would be less likely to engage in response distortion (Levashina et al., 2009). Therefore, more research is needed to understand the characteristics of individuals who endorse bogus items. Having a better understanding of who is likely to fake can better inform researchers and practitioners about how to address faking in the selection process, as well as how job outcomes may be affected.

Limitations

The current study is not without limitations. First, the voluntary research packet that contained the biodata inventory and bogus item measure informed participants that their responses would not affect their chances of progressing in the PMF selection process. Previous literature has also indicated that response inflation is more pronounced and self-report measures are less predictive when applicants are aware that their responses will impact selection decisions affecting potential employment (Anderson et al., 1984; Stark et al., 2001). Given that applicants knew the packet was voluntary, this may have diminished some of the hypothesized relationships.

Additionally, informing the participants that their responses were voluntary may also have contributed to the lower alphas that were observed for the biodata inventory

and bogus item scale. These measures did not quite reach the generally accepted .70 alpha level, and fell short of the recommended .90 level for applied settings (Nunally, 1967)[4]. It is possible that if the participants believed their responses on these measures would impact their progression in the PMF selection process, an increased consistency in these scales might have occurred. This may have provided a more accurate depiction of the true impact bogus item endorsement has on the biodata-AC task performance criterion-related validities, as well as increased the interpretability of the results.

Another limitation of the current study is also the focus on only one job-related outcome: task performance. Although the findings in the current study do not provide support for the notion that applicant faking impacts task performance, there are a variety of other job-related outcomes organizations are concerned about when making selection decisions. For instance, Griffith et al. (2007) suggested that the applicant faking may impact contextual behavior and Lewicki et al. (1997) proposed that faking may impact counter productive work behaviors (CWBs). Therefore, before we can be confident that faking has no adverse consequences on validity or job performance, more research is needed to better understand the potential impact applicant faking may have on a variety of job-related outcomes.

Finally, the characteristics of the PMF applicant pool may also have limited the generalizability of the current findings. As previously noted, all PMF nominees must demonstrate a desire to work for the Federal government and must obtain or be in the process of completing a graduate degree. Also, all participants in the current study had already passed a previous assessment in the selection process to be eligible to participate

---

[4]Confirmatory factor analyses were conducted to determine whether there was empirical justification for removing any items from the biodata or bogus item scales. Findings did not justify the removal of any items.

in the AC. Therefore, it is likely that restriction in range occurred within the current

applicant sample (Nunnally, 1967). This may have affected the biodata and AC

performance scores by reducing the amount of variance in the scores. Range restriction

could not be corrected in the current study because the biodata inventory was not

implemented as part of the first assessment phase; therefore, predictor scores were not

collected for the entire applicant pool. Unfortunately, this limitation was a function of

the multi-hurdle selection process that is driven by Federal policy, which did not take into

consideration the current study goals and the subsequent effects on future research

findings.

Future Research

The results of this study identify several areas of focus for future research. First,

additional research utilizing of the bogus knowledge scale is needed. As noted

previously, research using bogus knowledge items is relatively limited within the faking

literature. Further, of the studies that have utilized bogus item scales, there has not been

a consistent approach to categorize varying levels of response distortion. Levashina et al.

(2009) and Carroll (2007) chose to dichotomize responses to bogus items (e.g., faked or

honest response) and then sum the endorsed items; however, each study defined a faked

response differently. Levashina et al. considered a bogus item endorsement at any degree

a faked response, whereas Carroll defined a faked response as a bogus item endorsement

at a high magnitude (e.g., a response of 3 or 4 response on a 0-4 point scale). On the

other hand, Dwight and Donovan (2003) kept the bogus item scales continuous, summed

the endorsed bogus items, and then categorized respondents into three categories based

on their bogus item endorsements.

The current study used an approach similar to Dwight and Donovan (2003) in an attempt to maximize the variance in the bogus item scale and because this approach is most consistent with how the current study defines faking (e.g., faking magnitude as well as frequency of bogus item endorsement). However, it is worth noting that when applicants were categorized based on their bogus item score (i.e., Honest Respondent, Moderate Faker, and Extreme Faker), the effect of faking magnitude in predicting AC performance scores was lost. This would suggest that it may be more appropriate to operationalize bogus item scores as a continuous variable rather than grouping applicants into categories based on their score. Although this variable was categorized in the current study in order to observe differences in performance across levels of faking magnitude, additional research utilizing bogus knowledge scales is needed to better understand individual differences in response distortion and whether there is an optimal scoring approach.

Future research should also continue to explore the best way to utilize bogus item endorsement to score and classify varying levels of response distortion. The current study incorporated faking magnitude and variability using all seven bogus items. It is worth noting that two of the seven bogus items (bogus items 1 and 2) were endorsed at a higher frequency and magnitude than the others. Given that the majority of applicants indicated familiarity with these items, it is possible that these applicants may have endorsed them mistakenly[5]. If this is the case, these individuals' endorsements may be misinterpreted as intentional response distortion, which may impact how bogus items are

---

[5] Further inspection of these items indicated that they contained wording similar to real legislation, which may have contributed to the high degree of endorsements.

scored and interpreted[6]. Therefore, future research should consider protocols or procedures to ensure and enhance accuracy of the bogus item scales.

Additionally, the current study offers a better understanding of applicant faking behaviors in a real-world context and provides further support for the bogus knowledge scale as a valid measure of assessing individual differences in faking. Specifically, applicants varied with regard to the frequency and magnitude with which they endorsed bogus items in an applied setting. Approximately 91% of applicants exhibited some degree of response distortion by endorsing at least one bogus item. Further, the use of a bogus knowledge scale in the current study provides evidence of individual differences in response distortion. Nine percent of applicants did not endorse any bogus items, compared to 4% of applicants who endorsed all seven bogus items. These endorsement rates are in a similar pattern as previous research, with few individuals endorsing bogus items at high frequencies and magnitudes (Dwight & Donovan, 2003; Levashina et al., 2009). Some research has suggested that as many as half of an applicant sample will engage in some level of response distortion (Donovan et al., 2003; Fluckinger et al., 2008; Griffith et al., 2007). Given that the findings of the current study reported higher rates of response distortion, it may be worthwhile for future research to continue to utilize bogus knowledge scales to obtain a more accurate representation of response distortion in applicant settings.

While the bogus knowledge scale has demonstrated promise in identifying fakers, prior to implementation, practitioners may want to consider how applicants will respond

---

[6] Post-hoc analyses were conducted to recalculate bogus item score without the first two bogus items. This change affected the size of the response groups (the Honest Response Group doubled in size and the group sizes for the other two response groups decreased) and decreased the reliability of the bogus knowledge scale ($r = .56$). Despite these effects, the significance of the hypotheses was not affected.

to the use of this type of scale. If applicants recognize the item content as overly bogus

or learn that they were not selected for a position because they were distorting their

responses, it could create an undesirable impression or reputation about the organization.

Therefore, organizations should exercise caution in using bogus item scales and/or

consider using them with a disclaimer informing applicants that the selection process

contains means to detect dishonest responding. This is particularly important given how

easy it can be to incorporate bogus item scales into an existing selection system (e.g.,

they can be administered in conjunction with any non-cognitive assessment). This may

mitigate response inflation (Dwight & Donovan, 2003; Vasilopoulos, Cucina, &

McElreath, 2005) and abate potential negative applicant reactions to the scales.

Another consideration for future researchers and practitioners is to explore

administering bogus items in various methods (e.g., response latencies). Capturing

responses to bogus items in "real time" may provide additional information about

underlying faking processes. Specifically, using response latencies may help extend

existing researcher by providing insight as to whether fakers take more or less time to

respond to items than honest respondents (Robie, Brown & Beatty, 2007; Vasilopoulos,

et al., 2000; Vasilopoulos et al., 2005). Collecting this additional piece of information

may be useful in enhancing the ability to identify individuals who engage in response

distortion. Additionally, as the use of computers in the selection process continues to

grow, this may become an increasingly attractive option for organizations.

Future research should also continue to examine faking from a theoretical

perspective. Paulhus' model of Social Desirability (Paulhus, 1984) is one framework that

has been used to understanding faking. However, it has been argued that this two-

construct model may obscure important variations in response distortion (Griffith, 2005). For instance, Levin and Zickar (2002) has suggested that faking behavior can be described as a continuum, and proposed four levels of faking. Additional research examining the underlying traits and motivations that contribute to faking may be helpful to researchers to obtain more agreement about how to operationalize and define faking (Sackett, 2011). Further, having a better understanding of faking from a theoretical perspective may provide insight as to the best methods for detecting response distortion. If it is the case that faking is more complex than Paulhus' two-factor model, using a method such as the bogus item scale, which provides a more sensitive assessment of response distortion, may be a more appropriate means of assessing applicant faking.

Another avenue for future researchers is to examine individual difference variables that may correlate with varying levels of response distortion. Although researchers have suggested a relationship between faking and certain personality traits, this has not been examined with bogus knowledge items. Given that bogus items are believed to be capturing true response distortion, it is possible that personality traits associated with honesty (e.g., low integrity, high Emotional Stability) would be most indicative of intentional response distortion. Alternatively, Ramsay et al., 2008 suggested that bogus item endorsement may be related to Openness, and Pauls and Crost (2005) claim that faking reflects a positive, adaptive ability. Essentially, if this research line of research is correct, it would suggest that individuals who endorse bogus items do so because they are attempting to be open and adaptable, not deceitful. Therefore, future research should continue to explore various characteristics associated with bogus item

endorsement to better understand its impact on predicting performance as well as subsequent job-related outcomes.

Another contribution of the current research was the examination of the relationship between applicant faking and task performance. Although there is theoretical justification to believe that faking will impact subsequent job performance, no performance differences were found within the current sample. While it may be the case that there are no real differences in task performance across individuals who engage in varying levels of response distortion, there are still several factors future researchers should consider before coming to this conclusion. More specifically, future research would benefit from examining faking performance over an extended period of time. It is possible that applicants who fake in a selection context will then perform to the best of their ability once on the job. However, it is equally plausible that individuals who do not engage in response distortion in the selection context may go on to engage in dishonest behaviors once on the job. Currently, there is insufficient research to support either position; therefore, future research should explore whether faking has a long-term impact on subsequent job performance.

While previous researchers have suggested that faking may affect job performance positively and negatively, the findings in the current study suggest that faking had minimal impact on the predictor-task performance relationship. Despite the lack of support for the notion that there are performance differences between fakers and honest respondents, there is still reason to suspect that the true relationship between faking and performance is not as fruitless as the current findings suggest. For instance, Komar et al. (2008) suggested that the relationship between applicant faking and job

performance may depend on whether faking is needed to be successful in the job. In this sense, faking variance may be more or less useful in selection processes depending on the job in question. Therefore, more research utilizing various job outcome data is needed to identify other variables that may impact the faking-job performance relationship.

Having a better understanding of the applicant faking-performance would also be useful for determining how the bogus knowledge scale could be used as part of a selection process. Many organizations use selection procedures that select applicants based on a rank ordering of scores on a predictor measure. Because applicants inflate their responses on these self-report measures, there is a high risk of selecting individuals who inflated their responses. Understanding how applicant faking may be related to certain job-related outcomes could provide guidance in decisions about how to address applicants who engage in extreme faking magnitudes in selection settings. For instance, if faking is found to be related to negative job outcomes, using a bogus knowledge scale early in the selection process may be cost-effective for organizations, as individuals who are removed from the applicant pool based on amount of bogus knowledge endorsement would not participate in subsequent assessments that may be more costly. Additionally, using a bogus knowledge scale to identify applicants who inflate their responses could be useful in determining whether they should be removed from the selection process in order to make the process fairer for the applicants who do not engage in response distortion. Therefore, the current study is valuable to researchers and practitioners because it seeks to provide further insight about the extent to which faking poses a threat to selection processes utilizing non-cognitive measures and subsequent task performance.

Conclusion

The current study utilized a bogus knowledge scale to investigate the extent to which individual differences in faking attenuates the criterion-related validity of non-cognitive measures in predicting AC task performance. Although results for this hypothesis were not entirely supportive, more research is needed to better understand the impact faking has on non-cognitive measures in selection contexts. Additionally, this study examined whether there are differences in AC task performance scores among individuals who distort their responses on a biodata inventory relative to those who respond honestly. There was also little support for the notion that there are differences in AC task performance among individuals who engage in faking behaviors and those who do not.

In other words, these results suggest that if applicant Mary provides inaccurate information about herself when applying for a job, and is able to successfully raise her scores on selection measures in the process, her chances of being selected may also increase (because applicants with who endorsed more bogus items also tended to have higher biodata scores). While the current findings provide little support for the notion that whether Mary distorts her responses will impact her task performance once on the job, it is still unknown whether her response distortion would affect other aspects of her work performance.

It is unknown whether the lack of supported hypotheses is due to unreliable measures or if applicant faking is truly not a concern in selection settings (Ones et al., 1996). Therefore, future researchers should continue to use bogus knowledge items to assess individual differences in faking to better understand the impact faking has on

predicting performance. This research should also expand beyond task performance, to

include other important job-related outcome variables.

Table 1.

Descriptive Statistics for All Study Variables

	M	*SD*	*1.*	*2.*	*3.*
1. Biodata Inventory[1]	90.94	5.69	(.66)		
2. Assessment Center (AC)[2]	17.88	2.54	.12**	--	
3. Bogus Item Scale[3]	4.63	3.64	.16**	-.06*	(.66)

Notes: N=970. Cronbach's α coefficients are in parentheses. ** $p < .01$; * $p < .05$. [1]Biodata Inventory mean based on a total score of 105. [2]AC mean based on a total score of 25. [3]Bogus items were measured with a 5-point rating scale; the Bogus Item Scale mean is based on a total score of 28.

Table 2.

Results of Analysis of Variance with Biodata Score Across Faking Groups

| Response Group | N | Biodata Score | | d_1 | d_2 |
		M	SD		
Honest Response Group	86	90.69	5.51		
Moderate Faking Group	748	90.62	5.65	.01	
Extreme Faking Group	136	92.81*	5.72	-.38	-.39

Notes: *Mean is statistically different ($p < .05$) from the Honest Group. Bogus items were not included in the Biodata Inventory score. d_1 indicates standardized mean difference scores from the Honest Group. d_2 indicates standard mean difference scores from the Moderate Faking Group.

Table 3.

*Results of Multiple Regression Analyses with Biodata Score and Faking Group
(Categorical Variable)*

	b	SE b	β
Step 1			
Constant	17.88	0.08	
Biodata Score	0.05	0.01	0.12**
Step 2			
Constant	17.49	0.22	
Biodata Score	0.06	0.01	0.13**
Honest Response Group	0.43	0.35	0.05
Moderate Faking Group	0.46	0.24	0.08
Step 3			
Constant	17.51	0.23	
Biodata Score	0.05	0.04	0.10
Honest Response Group	0.41	0.36	0.05
Moderate Faking Group	0.44	0.25	0.07
Interaction with Biodata and Honest Response Group	-0.07	0.06	-0.01
Interaction with Biodata and Moderate Faking Group	0.02	0.04	0.03

Notes: Faking Response Group was represented as two dummy variables, with Extreme Faking Group as the reference group. $R^2 = .02$ for Step 1; $\Delta R^2 = .02$ for Step 2 ($p > .05$).; $\Delta R^2 = .01$ for Step 3 ($p > .05$). ** $p < .01$; * $p < .05$.

Table 4.

Results of Multiple Regression Analyses with Biodata Score and Bogus Item Endorsement (Continuous Variable)

	b	$SE\ b$	β
Step 1			
Constant	17.88	0.08	
Biodata Score	0.05	0.01	0.12**
Step 2			
Constant	17.88	0.08	
Biodata Score	0.06	0.01	0.13**
Bogus Item Endorsement	- 0.06	0.02	-0.09**
Step 3			
Constant	17.89	0.08	
Biodata Score	0.06	0.01	0.14**
Bogus Item Endorsement	-0.06	0.02	-0.08*
Interaction with Biodata and Bogus Item Endorsement	0.00	0.00	-0.01

Notes: R^2 = .02 for Step 1; ΔR^2 = .02 for Step 2 ($p < .05$).; ΔR^2 = .02 for Step 3 ($p > .05$). ** $p < .01$; * $p < .05$.

68

Table 5.

Results of Analysis of Variance with Mean Assessment Center Score Across Faking
Groups in the Top of the Biodata Distribution

		AC Performance			
Response Group	*N*	*M*	*SD*	*d₁*	*d₂*
Honest Response Group	23	17.83	2.37		
Moderate Faking Group	193	18.40	2.54	-.23	
Extreme Faking Group	56	17.97	2.80	-.05	.16

Notes: Means are not statistically different across groups. d_1 indicates standardized mean difference scores from the Honest Group. d_2 indicates standard mean difference scores from the Moderate Faking Group.

References

Abrahams, N.M., Neumann, I., Githens, W.H. (1971). Faking vocational interests: Simulated versus real life motivation. *Personnel Psychology, 24(1)*, 5-12.

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage Publications.

Alliger, G.M., Lillianfield, S.O., & Mitchell, K.E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7,* 32-29.

Anderson, C.D., Warner, J.L., & Spencer, C.C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69,* 574-580.

Arthur, W., Day, E.A., McNelly, T.L., & Edens, P.S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56,* 125-154.

Austin, E.J., Deary, I.J., Whiteman, M.C., Fowkes, F.G.R., Pedersen, N.L., Rabbitt, P., Bent, N. & McInnes, L. (2002). Relationships between ability and personality: Does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences, 32,* 1391-1411.

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

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

Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self and peer reports. *European Journal of Personality, 3,* 31-45.

Borgatta, E.F. (1964). The structure of personality characteristics. *Behavior Science, 12,* 8-17.

Bowler, M.C., & Woehr, D.J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91(5),* 1114-1124.

Bradley, K.M., O'Shea, P.G., & Hauenstien, N.M. (2002, April). *Factors related to personality test response processes and response endorsements.* Paper presented at the 17th annual meeting of the Society for Industrial and Organizational Psychology, Toronto, ON.

Carroll, S.A. (2007). *Examining dishonesty in the job application process: Relationships with social desirability, personality, and deviant behavior.* Unpublished doctoral dissertation, University of Calgary, Calgary, Alberta.

Christiansen, N.D. (1998, April). *Sensitive or senseless? On the use of social desirability measures to support inferences regarding applicant response distortion.* Paper presented at the 14th annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Cohen, Jacob. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Conley, J.J. (1985). Longitudinal stability of personality traits: A multitrait-multimethod-multioccasion analysis. *Journal of Personality and Social Psychology, 49(5),* 1266-1282.

Converse, P.D., Oswald, F.L., Imus, A., Hendricks, C., Roy, R., & Butera, H. (2009). Comparing personality test formats and warnings: Effects on criterion-related validity and test-takers reactions. *International Journal of Selection and Assessment, 16(2),* 155-169.

Costa, P.T., & McCrae, R.R. (1989). *The NEO PI/FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando, FL: Harcourt.

Crowne, D.P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Cunningham, M.R., Wong, D.T., & Barbee, A.P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology, 79,* 643-658.

Day, A.L., & Carroll, S.A. (2008). Faking emotional intelligence (EI): Comparing response distortion on ability and trait-based EI measures. *Journal of Organizational Behavior, 29(6),* 761-784.

Dilchert, S., Ones, D.S., Davis, R.D., & Rostow, C.D. (2007). Cognitive ability predicts objectively measured counter-productive work behaviors. *Journal of Applied Psychology, 92,* 616-627.

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.

Douglas, E.F., McDaniel, M.A., & Snell, A.F. (1996, August). The validity of non-

cognitive measures decays when applicants fake. *Proceedings of the Academy of*

Management.

Drasgow, F., & Kang, T. (1984). Statistical power of differential validity and differential

prediction analyses for detecting measurement nonequivalence. *Journal of*

Applied Psychology, 69(3), 498-508.

Dunnette, M.D., McCartney, J., Carlson, H.C., & Kirchner, W.K. (1962). A study of

faking behavior on a forced-choice self-description checklist. *Personnel*

Psychology, 15, 13-24.

Dwight, S.A., & Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human*

Performance, 16(1), 1-23.

Edwards, A.L. (1957). *The social desirability variable in personality assessment and*

research. New York: Dryden Press.

Egan, V. (1989). Links between personality and, ability and attitudes in a low-IQ sample.

Personality and Individual Differences, 10, 997-1001.

Ellingson, J.E., Sackett, P.R., & Hough, L.M. (1999). Social desirability corrections in

personality measurement: Issues of applicant comparison and construct validity.

Journal of Applied Psychology, 84, 155-166.

Eysenck, H.J. (1971). Relation between intelligence and personality. *Perceptual and*

Motor Skills, 32, 637-638.

Fluckinger, C.D., McDaniel, M.A., & Whetzel, D.L. (2008). Review of faking in

personnel selection. In Mandal, M. (Ed). *In search of the right personnel.* New

Delhi: McMillian.

Garry, R. (1953). Individual differences in ability to fake vocational interests. *Journal of Applied Psychology, 37(1),* 33-37.

Gatewood, R.D., & Field, H.S. (2001). *Human resource selection, 5th ed.* Orlando, FL: Harcourt.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72(3),* 493-511.

Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. *Personality and Individual Differences, 36(7),* 1545-1558.

Griffith, R. (1998). *Faking of noncognitive selection devices: Red herring is hard to swallow.* Unpublished doctoral dissertation, University of Akron, Akron, OH.

Griffith, R., Chmielowski, T.S., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36,* 341-355.

Griffith, R.L., Chmielowski, T.S., Snell, A.F., Frei, R.L., & McDaniel, M.A. (2000, April). *Does faking matter? An examination of rank order changes in applicant data.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology. New Orleans, LA.

Griffith, R.L., Rutkowski, K.A., Gujar, A., Yoshita, Y., & Steelman, L.A. (2005). *Modeling applicant faking behavior: New methods to examine an old problem.* Unpublished manuscript.

Grubb, W.L., & McDaniel, M.A. (2007). The fakability of Bar-On's Emotional Quotient Inventory short form: Catch me if you can. *Human Performance, 20,* 43-59.

Hakel, M.D. (1974). Normative personality factors recovered from ratings of personality

    descriptors: The beholder's eye. *Personnel Psychology, 27,* 409-421.

Hakstian, A.R., & Ng, E. (2005). Employment-related motivation distortion: Its nature,

    measurement, and reduction. *Educational and Psychological Measurement, 65,*

    405-441.

Helmes, E., & Holden, R.R. (1986). Response styles and faking on the Basic Personality

    Inventory. *Journal of Consulting and Clinical Psychology, 54,* 853-859.

Hogan, R. (1991). Personality and personality assessment. In M.D. Dunnette & L.M.

    Hough (Eds.), *Handbook of industrial organizational psychology* (pp. 873-919).

    Palo Alto, CA: Consulting Psychologists Press.

Hogan, R. (2006). *Personality and the fate of organizations.* Mahwah, NJ: Erlhaum.

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

    employment selection. *Journal of Applied Psychology, 92(5),* 1270-1285.

Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and

    employment decisions: Questions and answers. *American Psychologist, 51,* 469-

    477.

Hogan, R., & Nicholson, R.A. (1988). The meaning of personality test scores. *American*

    Psychologist, 43, 621-626.

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

    personality assessment. *Journal of Clinical Psychology, 37(2),* 379-386.

Hough, L.M. (1992). The 'Big Five' personality variables – construct confusion:

    Description versus prediction. *Human Performance, 5,*139-155.

Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 109-244.

Hough, L.M. (1998a). Personality at work: Issues and Evidence. In M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 131-166). Hillsdale, NJ: Erlbaum.

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*, 581-595.

Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 73-98.

Hurd, J.M., Barrett, G.V., Miguel, R.F., Tan, J.A., & Lueke, S.G. (2001, April). *When do response distortion scales reflect faking? A meta-analysis.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hurtz, G.M., & Donovan, J.J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869-879.

Huws, N., Reddy, P.A., & Talcott, J.B. (2009). The effects of faking on non-cognitive predictors of academic performance in university students. *Learning and Individual Differences, 19,* 476-480.

John, O.P. (1989). Towards a taxonomy of personality descriptors. In Buss D.M., Cantor

    N. (Eds.), *Personality psychology: Recent trends and emerging directions*. New

    York: Springer-Verlag.

Kim, B.H. (2006). *Building a strategic theory of deception.* Unpublished doctoral

    dissertation, Michigan State University, East Lansing.

Komar, S., Douglas, J.B., 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 tests: The use of idiosyncratic item responses. *International*

    Journal of Selection and Assessment, 15, 220-231.

Law, K.S., Mobley, & Wong. (2002). Impression management and faking in biodata

    scores among Chinese job-seekers. *Asia Pacific Journal of Management, 19,* 541-

    556.

Lautenschlager, G.J. (1994). Accuracy and faking of background data. In: Stokes, G.A.,

    Mumford, M.D. and Owens, W.A. (Eds), *Biodata Handbook*. Palo Alto, CA:

    Consulting Psychologists Press, pp. 391-419.

Levin, R.A. (1995, May). *Self-presentation, lies and bullshit: The impact of impression*

    management on employee selection. Presented at the annual meeting of Society of

    Industrial and Organizational Psychology, Orlando, FL.

Levashina, J., Morgeson, F.P., Campion, M.A. (2009). They don't do it often, but they do

    it well: Exploring the relationship between applicant mental abilities and faking.

    International Journal of Selection and Assessment, Vol 17(3), 271-281.

Levin, R.A., & Zickar, M.J. (2002). Investigating self-presentation, lies, and bullshit:

    Understanding faking and its effects on selection decisions using theory, filed,

    research, and simulation. In J.M. Brett & F. Drasgow (Eds.), *The psychology of*

    work: Theoretically based empirical research (pp. 253-276). Mahwah, NJ:

    Lawrence Erlbaum Associates.

Lewicki, R.J., Poland, T., Minton, J.W., & Sheppard, B.H. (1997). Dishonesty as

    deviance: A typology of workplace dishonesty and contributing factors. In R.J.

    Lewicki, R.J. Bies, et al. (Eds.) *Research on negotiation in organizations, 6,* 53-

    86).

Lorr, M., & Manning, T.T. (1978). Higher-order personality factors of the ISI.

    Multivariate Behavioral Research, 13, 3-7.

Lueke, S.B., Snell, A.F., Illingworth, A.J., & Paidas, S.M. (2001, April). *An empirical*

    test of an interactional model of faking. Paper presented at the 2001 meeting of

    the Society for Industrial and Organizational Psychology. San Diego, CA.

McCrae, R.R., & Costa, P.T. (1983). Social desirability scales: More substance than

    style. *Journal of Consulting and Clinical Psychology, 51,* 882-888.

McCrae, R.R., & Costa, P.T. (1985). Updating Norman's "adequate taxonomy":

    Intelligence and personality dimensions in natural language and in questionnaires.

    Journal of Personality and Social Psychology, 49, 710-721.

McFarland, L.A., & Ryan, A.M. (2000). Variance in faking across noncognitive

    measures. *Journal of Applied Psychology, 85*, 812-821.

McFarland, L.A., & Ryan, A.M. (2001). *Toward an integrated model of faking behavior.*
Paper presented at the 16th annual meeting of the Society for Industrial and
Organizational Psychology, San Diego, CA.

McFarland, L.A., Yun, G.J., Harold, C.M., Viera, L., & Moore, L.G. (2005). An
examination of impression management use and effectiveness across assessment
center exercises: The role of competency demands. *Personnel Psychology, 58,*
949-980.

Meriac, J.P. (2008). Further evidence for the validity of assessment center dimensions: A
meta-analysis of the incremental criterion-related validity of dimension ratings.
Journal of Applied Psychology, 93(5), 1042-1052.

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.

Mueller-Hanson, R.A., Heggested, E.D., & Thornton, G.C. (2006). Individual differences
in impression management: An exploration of the psychological processes
underlying faking. *Psychology Science, 48(3),* 288-312.

Nguyen, N. (2001). *Faking in situational judgment tests: An empirical investigation of
the work judgment survey.* Unpublished doctoral dissertation, Virginia
Commonwealth University, Richmond.

Noller, P., Law, H., & Comrey, A.L. (1987). Cattell, Comrey, and Eysenck personality
factors compared: More evidence for the five robust factors? *Journal of
Personality and Social Psychology, 53(4),* 775-782.

Norman, W.T. (1963). Toward an adequate taxonomy of personality attributes:

    Replicated factor structure in peer nomination personality ratings. *Journal of*

    Abnormal & Social Psychology, 66, 574-583.

Nunnally, J.C. (1967). Psychometric Theory. McGraw-Hill, NY.

Ones, D.S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused

    occupational personality scales (COPS) used in personnel selection. *International*

    Journal of Selection and Assessment, 9, 6-34.

Ones, D.S., Viswevaran, C., & Reiss, A.D. (1996). The role of social desirability in

    personality testing for personnel selection: the red herring. *Journal of Applied*

    Psychology, 81, 660-79.

Ones, D.S., Viswesvaran, C., & Schmidt, F. (1993). Comprehensive meta-analysis of

    integrity test validities: Findings and Implications for Personnel Selection and

    Theories of Job Performance. *Journal of Applied Psychology, 78(4)*, 679-703.

Pannone, R.D. (1984). Predicting test performance: A content valid approach to

    screening applicants. *Personnel Psychology, 37,* 507-514.

Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal*

    of Personality and Social Psychology, 46, 598-609.

Paulhus, D.L. (1986). Self-deception and impression management in test responses. In

    A. Angleitner & J.S. Wiggins (Eds.), *Personality assessment via questionnaire*

    (pp. 142-165). New York: Springer.

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R.

    Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social*

    psychological attitudes (pp. 17-59). New York: Academic Press.

Paulhus, D.L. (1998). *Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding – 7 user's manual.* Toronto, ON: Multi-Health Systems, Inc.

Paulhus, D.L. (2001). Socially desirable responding: The evolution of a construct. In H. Braun, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Hillsdale, NJ: Erlbaum.

Paulhus, D.L., Bruce, M.N., & Trapnell, P.D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21,* 100-108.

Paulhus, D.L., & Harms, P.D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence, 32,* 297-314.

Paulhus, D.L., Harms, P.D., Bruce, M.N., & Lysy, D.C. (2003). The over-claiming technique: Measuring bias independent of accuracy. *Journal of Personality and Social Psychology, 84,* 681-693.

Paulhus, D.L., & Reid, D.B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60(2),* 307-317.

Pauls, C.A., & Crost, N.W. (2005). Cognitive ability and self-reported efficacy of self-presentation predict faking on personality measures. *Journal of Individual Differences, 26,* 194-206.

Pulakos, E.D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9,* 241-258.

Ramsay, L.J., Kim, B.H., Oswald, F.L., Schmitt, N., & Gillespie, M.A., (2008, April). *Bogus items, their content, and responses under different situational constraints.* Paper presented at the 23rd annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Rosse, J.G., Levin, R., & Nowicki, M. (1999, April). *Does faking on personality tests really matter?* Paper presented at the 12th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Rosse, J.G., Stecher, M.D., Miller, J.L., & Levin, R.A. (1998). The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634-644.

Sackett, P.R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance, 24,*379-385.

Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82,* 30-43.

Schab, F. (1991). Schooling without learning: Thirty years of cheating in high school. *Adolescence, 26,*839-847.

Schmidt, F.L., & Hunter, J.E. (1992). Causal modeling of processes determining job performance. *Current Directions in Psychological Science,* 89-92.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

Schmidt, F.L., & Rothstein, H.R. (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction.* Stokes,

Garnett S. (Ed.); Mumford, Michael D. (Ed.); Owens, William A. (Ed.); pp. 237-260. Palo Alto, CA, US: CPP Books, 1994. xxii, 650 pp.

Schmidt, F.L., Shaffer, J.A., & Oh, I.-S. (2008). Increased accuracy of range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61(4),* 827-868.

Schmitt, M.J., & Ryan, A.M. (1993). The Big Five in personnel selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78,* 966-974.

Smith, G.M. (1967). Usefulness of peer ratings of personality in educational research. *Educational and Psychological Measurement, 27,* 967-984.

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

Snell, A.F., Sydell, E.J., & Lueke, S.B. (1999). Toward a theory of applicant faking: Integrating studies of deception. *Human Resources Management Review, 9,* 219-242.

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

Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Van Iddekinge, C.H., Raymark, P.H., & Roth, P.L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90,* 536-552.

Vasilopoulos, N.L., Cucina, J.M., & McElreath, J.M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology, 90,* 306-322.

Vasilopoulos, N.L., Reilly, R.R., & Leaman, J.A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology, 85(1),* 50-64.

Viswesvaran, C., Ones, D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81(5),* 557-574.

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

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

Zerbe, W.J., & Paulhus, D.L. (1987). Socially desirable responding in organizational behavior: A preconception. *Academy of Management Review, 12,* 250-264.

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

Zickar, M.J., Rosse, J.G., & Levin, R.A. (1996, May). Modeling of faking in selection context. In C.L. Hulin (Chair), *The third scientific discipline: Compulational modeling in organizational research.* The 11th annual meeting of the Society of Industrial Organizational Psychology, San Diego, CA.

Zickar, M.J., Rosse, J.G., Levin, R.A., & Hulin, C.L. (1996, April). *Modeling the effects of faking on personality tests.* Paper presented at the 11th annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Appendix A: Bogus Item Scale

PART D: KNOWLEDGE QUESTIONNAIRE

INSTRUCTIONS:

Part D contains questions concerning your knowledge of programs and activities of both national and international interest. Choose <u>one</u> answer from among the five alternatives presented by circling the letter corresponding to your answer. Answer each question or statement as accurately as you can based on YOUR experiences and not what you think you should say or what other people would say.

1. How familiar are you with the content of the *No Child Left Behind Act*?
    A. very familiar
    B. somewhat familiar
    C. familiar
    D. vaguely familiar
    E. I have never heard of the *No Child Left Behind Act*

2. How familiar are you with the activities of the *Sierra Club?*
    A. very familiar
    B. somewhat familiar
    C. familiar
    D. vaguely familiar
    E. I have never heard of the *Sierra Club*

3. **How familiar are you with the activities of the *Pan Arab Alliance for Peace in the Middle East*?[7]**
    A. very familiar
    B. familiar
    C. somewhat familiar
   D. vaguely familiar
    E. not at all familiar

4. **How familiar are you with the content of *Title 18 of the Equality in Education Act*?**
    A. very familiar
    B. familiar
    C. somewhat familiar
    D. vaguely familiar
    E. I have never heard of the *Equality in Education Act*

---

[7]Note: Bogus items are in bold.

5. How familiar are you with the activities of *Amnesty International*?

   A. very familiar
   B. somewhat familiar
   C. familiar
   D. vaguely familiar
   E. I have never heard of *Amnesty International.*

6. How familiar are you with the content of the proposed *Kyl-Feinstein Immigration Reform Act*?
   A. very familiar
   B. somewhat familiar
   C. familiar
   D. vaguely familiar
   E. I have never heard of Kyl-Feinstein Immigration Reform Act

7. How familiar are you with the content of the *Kyoto Accord*?
   A. very familiar
   B. somewhat familiar
   C. familiar
   D. vaguely familiar
   E. I have never heard of the *Kyoto Accord*

8. How familiar are you with the activities of the group *Citizens for Urban Crime Reduction*?
   A. very familiar
   B. somewhat familiar
   C. familiar
   D. vaguely familiar
   E. I have never heard of *Citizens for Urban Crime Reduction*

9. How familiar are you with activities of *OPEC*?
   A. very familiar
   B. familiar
   C. somewhat familiar
   D. vaguely familiar
   E. I have never heard of *OPEC*

10. How familiar are you with the content of the *Resident Protection Act*?
   A. very familiar
   B. somewhat familiar
   C. familiar
   D. vaguely familiar
   E. I have never heard of *Resident Protection Act*

11. On average, how many times per week do you watch the CNN Program *Larry King Live*?
    A. 1
    B. 2 to 3
    C. 4 to 5
    D. less than once a week
    E. I have never watched *Larry King Live*

12. **On average, how many times per week do you watch the PBS program *Bill Lee's Eye on the World*?**
    A. 1
    B. 2 to 3
    C. 4 to 5
    D. less than once a week
    E. I have never watched *Eye on the World*

13. On average, how many times per week do you watch the PBS program *Frontline*?
    A. 1
    B. 2 to 3
    C. 4 to 5
    D. less than once a week
    E. I have never watched *Frontline*

14. **On average, how many times a week do you watch the CNN program *Today's World*?**
    A. 1
    B. 2 to 3
    C. 4 to 5
    D. less than once a week
    E. I have never watched *Today's World*

15. On average, how many times per week do you watch the CNN program *Crossfire*?
    A. 1
    B. 2 to 3
    C. 4 to 5
    D. less than once a week
    E. I have never watched *Crossfire*

Figure 1. Illustration of Hypothesized Relationships for H2a
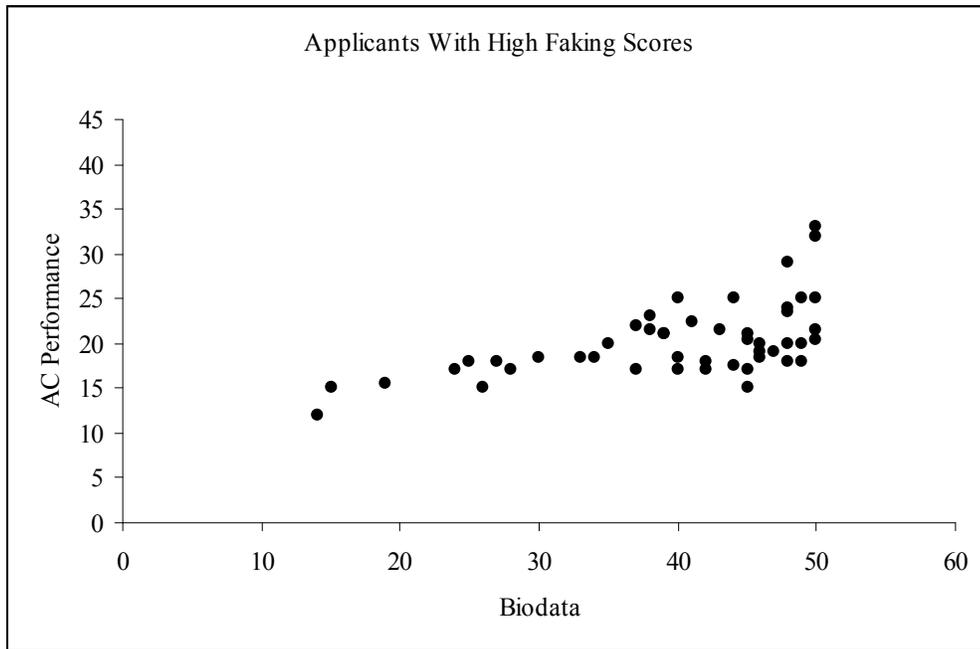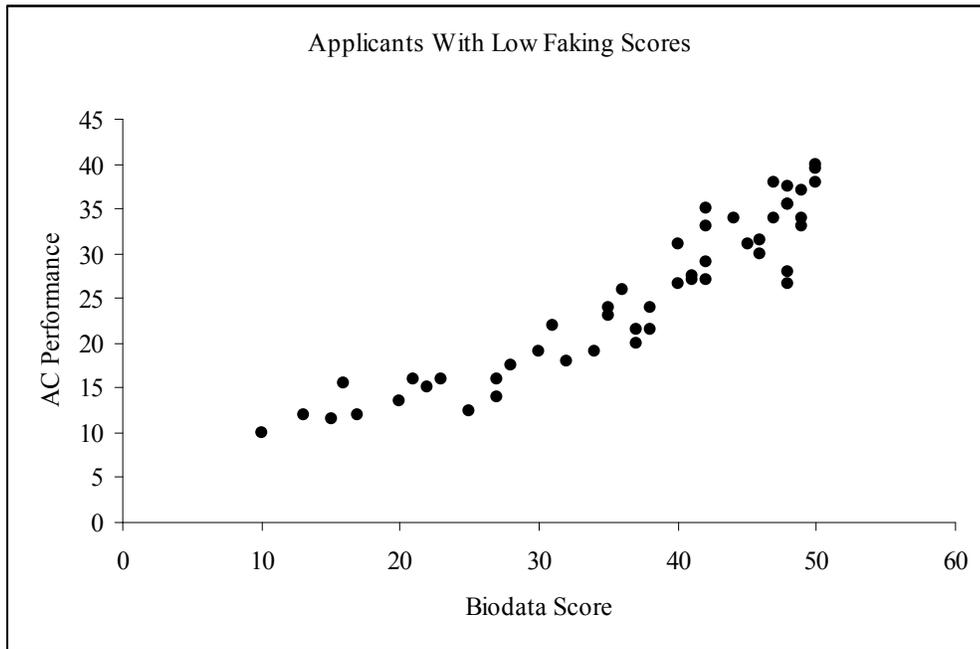
Note: These figures illustrate the hypothesized impact of applicant faking on predictive validity (as described on page 26); this is not actual data.

Figure 2. Illustration of Hypothesized Relationships for H2b
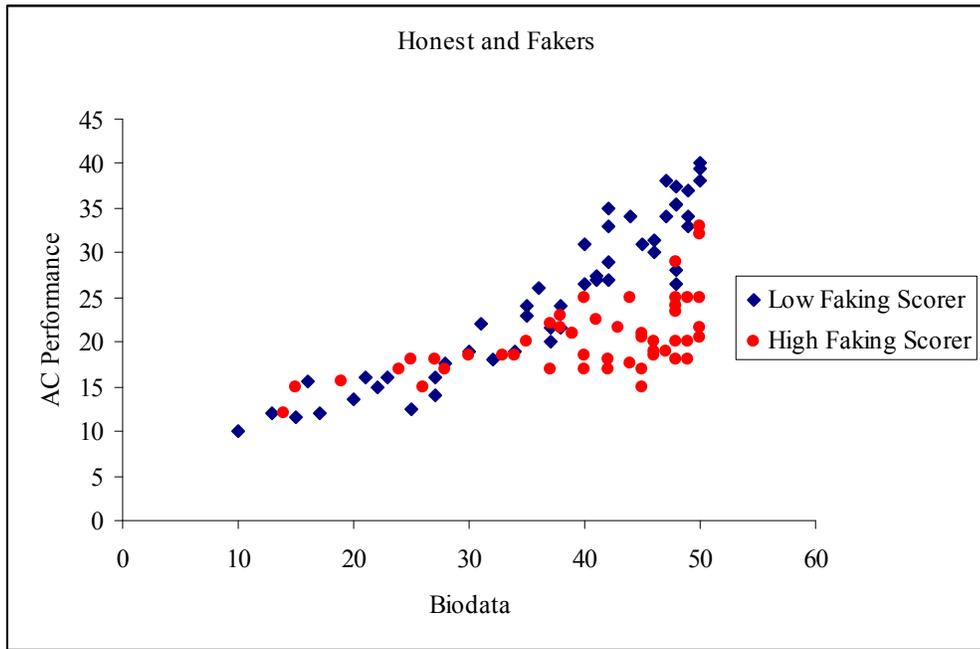
Note: These figures illustrate the hypothesized impact of applicant faking on predictive validity (as described on page 27); this is not actual data.