

**RESPONSE DISTORTION AND SOCIAL DESIRABILITY IN HIGH-
LEVEL EXECUTIVES**

A Thesis

Presented to

The Academic Faculty

by

Katherine A. Schnure

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in the

School of Psychology

Georgia Institute of Technology

December 2009

**RESPONSE DISTORTION AND SOCIAL DESIRABILITY IN HIGH-
LEVEL EXECUTIVES**

Approved by:

Dr. Lawrence James, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Charles Parsons
College of Management
Georgia Institute of Technology

Dr. James Roberts
School of Psychology
Georgia Institute of Technology

Date Approved: October 28, 2009

ACKNOWLEDGEMENTS

I would like to thank my committee members for their insightful commentary and guidance throughout this process. I would also like to express my gratitude to the members of the EA team at the consulting firm which collected the data used in this study for their willingness to share some of the fruits of their labor with me, and for their continued support in all areas of this project.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	ix
<u>CHAPTER</u>	
1 Introduction	1
Can People Distort Their Responses?	2
Why Do People Distort Their Responses?	6
Do Social Desirability/Impression Management Scales Detect Response Distortion?	12
Does Response Distortion Affect Validity?	14
2 Present Research Question and Hypotheses	16
Present Research Question	16
Primary Hypotheses	17
Exploratory Analyses	19
3 Method	21
Participants	21
Measure	23
4 Results	25
Hypothesis 1	25
Hypothesis 2	30
Hypothesis 3	30

Hypothesis 4	31
5 Discussion	33
Limitations	40
Future Directions	41
REFERENCES	42

LIST OF TABLES

	Page
Table 1: Univariate F-test Results	26
Table 2: 16PF Scale Means and Standard Deviations	27
Table 3: 16PF scale correlations with Impression Management (IM) scale	34
Table 4: Scale Correlations with Impression Management (IM) scale from norms	34

LIST OF FIGURES

	Page
Figure 1: Mueller-Hanson et al. (2006) Model of Faking	10

SUMMARY

The effect of Socially Desirable Responding (SDR) and response distortion on personality and selection measures has been debated for decades (Edwards, 1957b; Messick & Jackson, 1961; Morgeson et al., 2007, Ones, Viswesvaran, & Reiss, 1996). One area of this broad topic that has received less attention over the years has been the effectiveness of the scales used to measure SDR when evidence of potential response distortion is present. Using a MANOVA, this study found significant differences between job candidate and incumbent responses on the scales of the 16PF (Cattell & Cattell, 1995), which can be indicative of SDR (Rosse et al., 1998). However, no difference between the groups was found on the Impression Management scale used to measure SDR. Differences based on the job function of the respondents were also explored.

CHAPTER 1

INTRODUCTION

Personality tests and inventories have been a mainstay in industrial and organizational psychology for many years. One of the consistent criticisms of self-report personality measures used in selection and assessment processes is that the measures are designed in such a way that they can be easily faked, manipulated, or distorted. In fact, questions arose about potential problems with social desirability and faking on personality measures even before many of today's most popular measures were created (Edwards, 1957b; Messick & Jackson, 1961), and the debate rages on in the Industrial and Organizational Psychology research today.

As their use has continued to increase in the last decade (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006), questions about the extent to which the responses to the items on personality measures can be distorted have become more and more prevalent in the literature (Morgeson, Campion, Dipboye, Hollenbeck, Murphy & Schmitt, 2007a). The debate even spills into what the phenomenon should be called. Some use the terms "faking," "response distortion," "impression management," and "dissimulation" interchangeably, while others claim stark differences between the words. For the purposes of this research, the definitions used by Paulhus (2002) will be employed. Paulhus defines response biases as "any systematic tendency to answer questionnaire items on some basis that interferes with accurate self-reports (p.49)," and response sets as "short-lived biases attributable to some temporary distraction or motivation (p.49)." This research focuses on response bias, and more specifically, on

socially desirable responding, which is a type of response bias defined by Paulhus as “the tendency to give overly positive self-descriptions (p.50).” For consistency’s sake, the term response distortion will be used to delineate socially desirable responding, and “faking” as a concept.

The larger questions surrounding response distortion on personality measures often fall into one of four areas: 1) Whether people distort their responses to personality measures depending on their reason for taking the measure (e.g., applicant versus incumbent), 2) Why people may distort their responses on self-report personality measures, 3) Whether self-report measures are able to control for or detect response distortion and socially desirable responding using impression management or social desirability scales , and 4) Whether response distortion affects the validity of personality measures.

The research is perhaps the most numerous and contentious in response to the fourth question, the effect of response distortion and socially desirable responding (SDR) on the validity of personality measures. Though that is an interesting and rich area, it is only tangentially related to the purpose of this study. Therefore, this will only be visited briefly, and solely for the purpose of informing certain areas of the research objectives of this study. This study will focus on the first three questions related to response distortion.

Can people distort their responses on personality measures?

Issues with response distortion on personality measures have been argued for decades (Edwards, 1957a, 1957b; Guion & Gottier, 1965; Krug, 1978; Morgeson et al., 2007a). In looking back at some of the literature from personality’s first heyday in the

1940s and 1950s, we can see a reflection of the same arguments circulating in the research today. Messick and Jackson (Jackson, 1960; Messick & Jackson, 1961) made an early distinction between acquiescence, or the tendency to agree with statements, and social desirability. They looked at the effect of acquiescence on the California Personality Inventory (CPI; Jackson, 1960) and the Minnesota Multiphasic Personality Inventory (MMPI; Messick & Jackson, 1961). Using a factor analysis, they suggested that at the extreme, the two major factors found for the MMPI could be rotated into positions that could be interpreted as two response styles: the acquiescence and social desirability (Messick & Jackson, 1961, p. 302). This discovery led them to suggest further research into the development of a “refined interpretation of these and other stylistic consistencies (p. 303).” Their information on acquiescence was used in part to create early “bias” scales, measuring the tendency to blindly agree with items.

Additionally, Edwards (1957b) actually alluded to one of today’s most visible problems with self-report personality measures in a book he wrote in the late 1950s:

The potentially distorting influence of this tendency on scores based on subjects’ responses to statements in personality inventories has long been recognized by many psychologists. Yet, as late as 1946, Meehl and Hathaway (1946) could stress that, despite good reasons for believing in the existence of the tendency, very few systematic efforts had been directed toward doing anything about it. (p. vii)

This early research on response distortion looked at the problem of social desirability in self-report personality measures, which has been at the epicenter of the response distortion argument since its inception. The idea that people will tend to skew their

answers in a socially desirable direction so that they appear in a better light was recognized almost immediately. Early research found a product-moment correlation of .87 between the probability a certain item would be endorsed and the pre-determined social desirability scale value for that item when people had to sign their names on their answer sheet, thus removing anonymity (Edwards, 1957a). The study was replicated, with the measures taken anonymously the second time, and the correlation was still .83, suggesting that anonymity wasn't a factor in peoples' tendency to answer personality measure items in a socially desirable fashion.

Early findings by Edwards (1957a, 1957b), Meehl & Hathaway (1946), and Messick & Jackson (1961) may have led to a "revamping" of items on personality measures, and also played a part in the inclusion of social desirability scales, acquiescence scales, and more broad, simple "lie" scales in many of the tests. These scales were designed to alert test administrators to situations in which a test-taker may be endorsing items based on their social desirability, or what he/she felt the response "should" be given social and cultural norms, rather than endorsing items in a fashion that accurately reflected his/her personality. Most of the personality measures used today include at least one scale that looks at response distortion or social desirability that is taken into account during the interpretation of the test results (Cattell, Eber, & Tatsuoka, 1970).

After nearly 40 years of back-and-forth, today's research on response distortion rarely suggests outright that people are entirely unable to fake or distort their responses on personality measures (Birkeland et al, 2006; Hough et al, 1990; Morgeson et al, 2007a). In the last ten or so years, the most basic debate in the response distortion

argument has shifted ever so slightly from questioning whether or not response distortion is possible to determining whether or not people actually *do* fake their responses (Hogan, Barrett, & Hogan, 2007; Ones, Viswesvaran, & Reiss, 1996). This slight shift has not contributed greatly toward any resolution regarding whether or not response distortion a) occurs or b) has any effect on personality measures within the normal population.

There are those that believe all high-stakes or high-stress assessments will involve some level of participant deception (Morgeson et al, 2007a), and even the strongest proponents of self-report measures will concede that some people are capable of distorting their responses to a self-report in order to appear in a more positive light, especially when a job may be on the line (Dilchert, Ones, Viswesvaran, & Deller, 2006; Hogan, Barrett, & Hogan, 2007; Tett & Christiansen, 2007). In fact, those on both sides of the argument appear to support the evidence that people with certain attitudes and personality characteristics may be more prone to response distortion (Morgeson et al., 2007a; Rossé, Stecher, Miller, & Levin, 1998). Others lend credence to the idea that there may be a more defined model of response distortion based on a participant's ability to fake, with some engaging in self-deception, and others in impression management. Those engaged in self-deception are ignorant of their distortion, believing they are answering truthfully when, in fact, they are not. Those engaging in impression management are deliberately answering particular items in a way that they believe will cause them to appear attractive to an employer, and may or not be conscious of doing so (Mueller-Hanson, Heggstad, & Thornton, 2006; Paulhus, 1984).

Those supporting the notion that response distortion can and does occur on personality measures have conducted research that examines selection and assessment

situations where one group may have more motivation to fake than a second comparison group. One of the most commonly used approaches to measure response distortion has been to compare the scores of those completing a personality measure as job applicants against those taking the measure in a situation that will not have an effect on whether or not they obtain or keep a job (i.e. incumbents). Several studies have shown a significant difference between applicant and non-applicant scores on personality measures, suggesting applicants may be more apt to skew their answers in a socially desirable direction (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Hough, Eaton, Dunnette, & Kamp, 1990; Morgeson et al, 2007a, Rosse et al., 1998). It is thought that this may occur because applicants have more at stake than incumbents (Robie, Zickar, & Schmit, 2001; Rossé et al., 1998), as applicants are hoping to be able to get a job, while incumbents may not feel as great a need to impress those administering the measure.

Rossé, Stecher, Miller, & Levin (1998) contend that individuals are not only capable of faking on personality measures, but will fake their responses in a selection context. The authors found significantly greater levels of response distortion in job applicants versus job incumbent groups, and further argued that response distortion in job applicants has an impact on who is hired. There has been some support for this theory (Morgeson et al., 2007), but others have conducted research that shows response distortion has no adverse impact on selection (Bradley & Hauenstein, 2006; Ones et al., 2007; Ones, Viswesvaran, & Reiss, 1996).

Why do people distort their responses on self-report personality measures?

The theories that explain why people can or do fake or distort their responses on personality measures are more complicated than the relatively simple explanation for

why differences between applicants and incumbents emerge on tests. While the fact that applicant test-takers have more at stake in a selection situation than incumbents when completing a measure as a part of an assessment or basic experiment is easy to understand, the cognitive processes behind score differences are considerably more complex.

Paulhus (1984) outlined an early two-factor model of socially desirable responding. In this model, Socially Desirable Responding (SDR) occurred for one of two reasons: as part of “Self-Deception” or as part of “Impression Management.” Any unrealistically positive pattern about which the test-taker appeared to be wholly convinced was considered to be self-deception, which was regarded as an unconscious process. To describe the second factor in SDR, Paulhus arrived at the term “Impression Management” after revamping Sackeim and Gur’s (1978) concept of “other-deception.” Like many, Paulhus noted that continually feeling obliged to present a positive image of oneself could be regarded as a part of personality, rather than strictly as deception, which insinuates an intention to falsely represent oneself (Paulhus, 2002).

The question of “frame of reference” frequently comes into play as part of this aspect of response distortion as well, suggesting that applicants may respond to personality measure items using a different context than non-applicants (Schmit, Ryan, Stierwalt, & Powell, 1995). The applicant and non-applicant may have a different conditional disposition, which is the tendency to express behaviors that are contingent upon the situational conditions (Schmit et al, 1995, p.608). Because the job candidates may be taking the personality measure with the job for which they are under consideration in mind, they may be considering what they would do in that job in

choosing their responses. This frame-of-mind places greater emphasis on how a test-taker would react in the future given their vision of a role rather than how they would answer based on their prior experiences or their present state-of-mind. Their answers might differ from incumbents who may be responding to items with regard to the jobs they already have, referencing their experience in those jobs and the expectations those jobs entail (Smith, Hanges, & Dickson, 2001).

There have been several complex models the processes behind faking behavior or response distortion proposed over the years. Snell et al. (1999) presented an interactional model which integrated both contextual factors and constructs in individual differences that were commonly associated with faking or response distortion behavior. Those individual differences were placed into one of two categories: the ability to fake or distort, and the willingness to do so. Factors like general mental ability, the ease of faking on the particular measure (including item type and format), and knowledge of what is being measured by the test were characteristics included in the category of “Ability to fake” (Snell et al., 1999). “Willingness to fake” was determined by demographic factors, and additionally contextual factors (e.g. the presence or absence of warnings, the importance of test outcome, etc.) and dispositional factors such as integrity, and the tendency to be manipulative (Snell et al., 1999). Critics of this model point to the lack of exploration into personality factors that may also be associated with the “willingness” factor (Muller-Hanson, Heggstad, & Thornton, 2006).

A second model was proposed by McFarland and Ryan (2000, 2001), which was based largely upon Azjen’s (1991) theory of planned behavior. This model asserts that faking behavior is a result of intention to fake, and that intention to fake is born out of a

mixture of attitudes toward faking, subjective norms, and perceived behavioral control. According to this model, attitudes toward faking were influenced by the perceived rightness and wrongness of faking, subjective norms were influenced by beliefs about how others felt about faking, and perceived behavioral control was determined by beliefs about the ease or difficulty of faking (McFarland & Ryan, 2000). Though the model was supported by data (McFarland & Ryan, 2001), the model lacks Snell et al.'s (1999) dispositional factors that can include factors like the ever-important applicant status, and other elements central to the “faking” arguments.

Additional research into the processes involved with faking and response distortion focuses on individual differences (Mueller-Hanson, Heggstad, & Thornton, 2006), and provides one of the newest models of faking behavior. This model integrates those proposed by Snell et al. (1999) and McFarland & Ryan (2000, 2001), and additionally incorporates specific personality factors into the model. As shown in Figure 1, the hypothetical model theorized that faking behavior was again preceded by intention to fake, and that intention to fake was preceded by a number of antecedents: perceptions of the situation, ability to fake, willingness to fake, and “Big Five” personality factors Conscientiousness and Emotional Stability (Mueller-Hanson et al., 2006).

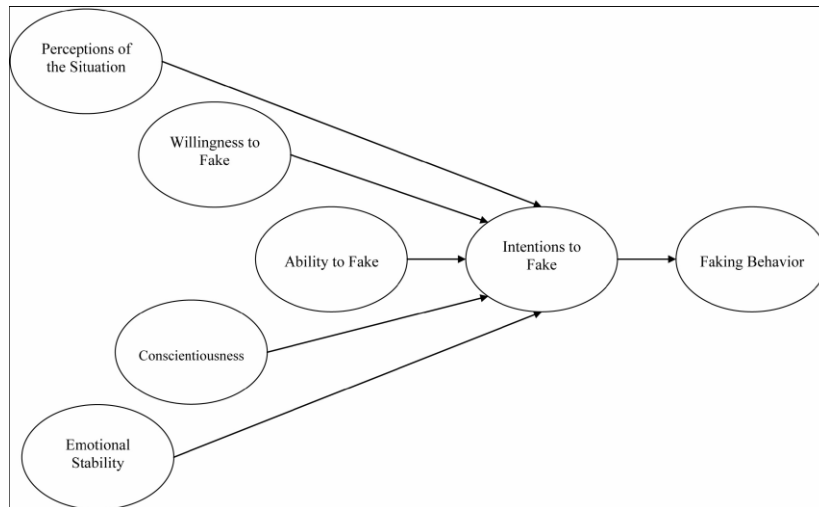


Figure 1 From Mueller-Hanson et al., 2006 p. 4

While the results supported the inclusion of Conscientiousness, Emotional Stability, and perceptions of the situation in the model, oddly, the data showed that ability to fake did not have a strong correlation with intention to fake, and willingness to fake showed a significant negative relationship with intention to fake. Overall, however, this model supported Paulhus's (1984, 2002) notion of faking or response distortion as an impression management strategy that can vary from person to person, rather than the earlier notion of deliberate deception put forth by Sackeim and Gur (1978) among others.

Additional explanations as to why people may fake their responses on self-report personality measures are still under examination. There are those that argue that this might be a difficult question to resolve, as researchers may be prone to using non-empirical evidence to back up their claims, which risks becoming what Dilchert and colleagues (2006) would call "pseudo-scientific lore" (p. 213), meaning that psychologists may use pure postulation to arrive at a conclusion about these cognitive processes.

Ones, Viswesvaran, and Reiss (1996) provide the strongest support for self-report personality measures regarding their “fakability” and validity. The authors are generally of the belief that faking doesn’t often occur in non-experimental settings, and dismiss the role of social desirability as a “red herring,” stating that it has no effect on the criterion-related validity of personality measures, and is not nearly as big of a problem in industrial and organizational psychology as past literature examining the topic has led us to believe. They also contend that social desirability scale scores reflect individual differences in personality constructs. More specifically, high scores on social desirability scales are related to emotional stability and conscientiousness (Ones, Viswesvaran, & Reiss, 1996). In their meta-analysis of response distortion and social desirability research, they argue that “removing the effect of social desirability from the Big Five dimensions of personality does not improve criterion-related validity of personality constructs (p.670).”

Barrick and Mount (1996) examined both factors in Paulhus’s (1984) model of socially desirable responding, looking at response distortion related to both self-deception and impression management, and their potential effects on the predictive validity of the personality constructs. Unlike the aforementioned study, Barrick & Mount noted that applicants did, in fact, engage in response distortion in the form of both self-deception and impression management. However, their research supported the claim that the distortion did not have an effect on the predictive validity of the personality constructs. Additional research done by others provides further support for these claims (Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough, 1998; Ones et al., 2007).

Holden (2008) found significant effects for response distortion on test validity using two different personality measures on two different samples, and suggests that

other studies may be underestimating the extent to which induced-faking research can be generalized to non-experimental samples. More specifically, the author notes that the use of moderated regression in tandem with the use of socially desirable responding scales causes the faking effect sizes to be more severely underestimated than many would believe (Holden, 2008, p. 320). Holden also notes that social desirability scales may be better measures of personality constructs or traits, as they are “fallible indicators of faking (p.320).”

Do Social Desirability/Impression Management Scales Detect Response Distortion?

In the extensive research done on socially desirable responding (SDR), one of the most common and logical methods of describing types of socially desirable responding is a two-factor method (Wiggins, 1964, Damarin & Messick, 1965; Sackeim & Gur, 1978; Paulhus, 2002). The two factors were perhaps most concisely described by Sackeim and Gur (1978) as “self-deception” and “other deception,” and later by Paulhus (1984) as “self-deception” and “impression management.” Self-deception refers to non-deliberate response distortion, where the test-taker may believe they are something they are not, while “impression management” allows for both deliberate lying and the “habitual presentation of a specific positive public impression (Paulhus, 2002, p. 56)” which may not be regarded as actual “deception.”

The first Social Desirability scale was created by Edwards (1957b), and was a 79-item scale that had showed perfect agreement in an initial sample that asked participants to respond in a socially desirable fashion. The chance probability of such agreement was less than .001 on each item. Often called impression management scales or social desirability scales, they tend to include items that the majority of people aiming to appear

positively would likely answer in one direction. In their most generic form, these scales measure the responses to statements like, “I have never lied,” and would consider those who answer “Correct” or “Agree” to such a statement as people who are likely answering in a socially desirable fashion, rather than honestly. While it would seem more impressive to an employer to suggest that you have never lied, and that you are a very honest and moral person, in actuality it is extraordinarily unlikely that the person taking the test has never told even so much as a white lie so long as he or she can remember. The individual that is interpreting a test with an abnormally high social desirability score would be advised to place less trust in the rest of the test-taker’s results, sometimes going so far as to suggest that the participant be administered another version of the test to see if the scores align appropriately (SHL Group, 1999).

That said, some research suggests that the interpretation of lie scales should be modified, as aiming to appear socially desirable may not be a negative attribute in a real-world selection situation (Elliott, 1981). As Edwards (1957b) stated during a time that saw the creation of many of the social desirability scales found in today’s measures, “It is obvious... that what is considered desirable and undesirable in the way of personality traits is culturally determined (p. 8).” It follows that if an item is viewed as highly desirable in a particular culture, its endorsement would also be viewed as positive by that culture. Possessing socially desirable attributes could certainly be a good thing in a selection situation, so it is important to draw the line between actually possessing socially desirable attributes, and endorsing items so as to appear more desirable in the eyes of test administrators.

Does Response Distortion Affect the Validity of Personality Measures?

As mentioned earlier, the concern about the effect of response distortion on the validity of personality measures is a heavily-researched area, but will be discussed only briefly in this review. There are those that contend that any response distortion that occurs has no effect on the validity of the measure (Barrick & Mount, 1996; Hogan, Barrett, & Hogan, 2007; Hough et al., 1990; Ones, Dilchert, Viswesvaran, & Judge, 2007; Ones, Viswesvaran, & Reiss, 1996), and then there are others who question that claim, suggesting that faking and response distortion have serious implications for the validity of personality measures, and that previous research could be underestimating the actual effect (Holden, 2008; Morgeson et al., 2007; Rossé, Stecher, Miller, & Levin, 1998).

Ones, Viswesvaran, and Reiss (1996) provide the strongest support for self-report personality measures regarding their “fakability” and validity. The authors are generally of the belief that faking doesn’t often occur in non-experimental settings, and dismiss the role of social desirability as a “red herring,” stating that it has no effect on the criterion-related validity of personality measures, and is not nearly as big of a problem in industrial and organizational psychology as past literature examining the topic has led us to believe. They also contend that social desirability scale scores reflect individual differences in personality constructs. More specifically, high scores on social desirability scales are related to emotional stability and conscientiousness (Ones, Viswesvaran, & Reiss, 1996). In their meta-analysis of response distortion and social desirability research, they argue that “removing the effect of social desirability from the Big Five dimensions of personality does not improve criterion-related validity of personality constructs (p.670).”

Barrick and Mount (1996) examined both factors in Paulhus's (1984) model of socially desirable responding, looking at response distortion related to both self-deception and impression management, and their potential effects on the predictive validity of the personality constructs. Unlike the aforementioned study, Barrick & Mount noted that applicants did, in fact, engage in response distortion in the form of both self-deception and impression management. However, their research supported the claim that the distortion did not have an effect on the predictive validity of the personality constructs. Additional research done by others provides further support for these claims (Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough, 1998; Ones et al., 2007).

Holden (2008) found significant effects for response distortion on test validity using two different personality measures on two different samples, and suggests that other studies may be underestimating the extent to which induced-faking research can be generalized to non-experimental samples. More specifically, the author notes that the use of moderated regression in tandem with the use of socially desirable responding scales causes the faking effect sizes to be more severely underestimated than many would believe (Holden, 2008, p. 320). Holden also notes that social desirability scales may be better measures of personality constructs or traits, as they are "fallible indicators of faking (p.320)."

CHAPTER 2

PRESENT RESEARCH QUESTION AND HYPOTHESES

Present Research Question

While many studies have set out to measure whether or not response distortion has an effect on the validity of self-report scales (Holden, 2008; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), far fewer have looked at the effectiveness of so-called “lie scales” (Elliott, 1981) and of social desirability or impression management scales. This study will look for differences in job candidate and incumbent scores on a personality measure that are consistent with patterns of socially desirable responding (SDR), and determine whether the response distortion scale (e.g. social desirability scales) in the personality measure is effectively reflecting that SDR may be taking place.

Although extensive research has been done looking at the scores of applicants versus non-applicants on personality measures (Rossé et al., 1998), and examined induced faking in undergraduate student subject studies (Mueller-Hanson, Heggstad, & Thornton, 2006), few studies have broken down the job function of subjects into relatively specific areas (Birkeland et al, 2006). It appears even fewer studies have looked specifically at the industry of the companies at which the subjects are employed or applying to be employed (Elliott, 1981), with those that do comparing few and seemingly random industries.

The present study will examine each of these constructs in relation to response distortion on the 5th edition of the Sixteen Personality Factor personality measure (16PF; Cattell & Cattell, 1995), which is a widely-used measure than can be employed in a selection context. Data will be procured from a large sample of high-level executives in

various roles across a variety of industries throughout the United States and Europe who have completed the 16PF as part of an executive assessment a) related to a succession management or talent review project, or b) as part of a selection procedure.

Prior research on the “fakability” of the 16PF has been conducted as part of the validation of various editions of the measure (Braun & Lafaro, 1968), and corrections for faking on the 16PF have been proposed (Christiansen, Goffin, Johnston, & Rothstein, 1994). However, the overall effectiveness of the Impression Management scale as a method for detecting response distortion, especially the comparison of candidate and incumbent samples, has not yet been examined in any great detail.

Primary Hypotheses

Hypothesis 1

Past studies have repeatedly shown a significant difference between applicant and non-applicant scores on personality measures, supporting the idea that applicants may be more apt to skew their answers in a socially desirable direction (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Hough, Eaton, Dunnette, & Kamp, 1990; Morgeson et al., 2007a, Rosse et al., 1998). This difference may be a result of the fact that applicants have more riding on the outcome of the personality measure than incumbents, and have more reason to make a good impression (Robie, Zickar, & Schmit, 2001; Rossé et al., 1998). Given these findings, *this study hypothesizes a statistically significant difference between candidate and incumbent groups on the 16PF*. Because prior research has not examined scale-by-scale differences on the 16PF specifically, it is unclear which of the specific scales will show differences, and exactly how many of the 16 scales will differ. However, if response distortion related to social desirability is driving the

difference between the groups, with the job candidates skewing their answers in a more socially desirable direction, it is expected that the individual 16PF scales that have strong correlations with the social desirability scale on the 16PF, Impression Management scale, will be more likely to show univariate scale differences. Specific scale differences can be examined using a post-hoc procedure if an overall difference is found.

Hypothesis 2

Social desirability scales exist in part to indicate whether a test-taker may be “attempting to consciously or unconsciously oversell themselves” (SHL Group, 1999, p. 68). If the results expected in H1 are found, the presence of significant mean differences on multiple scales of a personality measure could suggest that response distortion is taking place in one of the groups. If the differences in the scales are suggestive of socially desirable responding (SDR), that difference between the groups should also be reflected in the scale that is specifically designed to measure social desirability or impression management, which in the case of the 16PF is the Impression Management scale. On the contrary, *this study hypothesizes that the 16PF scale designed to measure SDR and response distortion will not show a significant difference between the candidate and incumbent groups.* This outcome would suggest that the social desirability/impression management scale may not be measuring faking and response distortion levels as well as it should, but could perhaps be measuring other characteristics.

Exploratory Analyses

Hypothesis 3

Though research has been done looking broadly at job function in relation to response distortion (Birkeland et al, 2006), a more detailed breakdown of job roles would be useful to help determine whether there is a difference in response distortion levels in different job functions. Some have suggested that jobs in fields like Sales & Marketing that require “positive” personality characteristics for success will be subjected to higher levels of response distortion than fields like Finance, where success is more heavily reliant on a particular skill than on personality (Birkeland et al, 2006; Morgeson et al, 2007a). *This study hypothesizes that there will be a significant difference in 16PF scores based on the job function of the applicants.* Due to the relative lack of research highlighting individual scale differences, which scales will differ and exactly how many will differ is unclear.

Hypothesis 4

In a 2007 *Personnel Psychology* article discussing the issues with self-report personality measures, Hollenbeck raised the example of working at Disney (Morgeson et al 2007): “If you work at Disney, you are supposed to have a good day every single day...if you cannot fake a personality item, how are you going to work at Disney everyday, smiling at these young children stomping on your feet?” (p. 717). Though Hollenbeck uses the example in reference to job function, the example may also apply to the industry of the company. For example, do those employed at a large, multi-national industrial manufacturing companies show different levels of “warmth” than those employed in the leisure and hospitality industry independent of job function? The

differences may be a result of self-selection, or may be induced by the nature of the industries themselves, but *the present study hypothesizes that there will be statistically significant differences in scale scores between industries*. Once again, the specific scale differences are not being hypothesized.

CHAPTER 3

METHOD

Participants

The 16PF personality scale was administered to approximately 1,750 executives between 2002 and 2009 as part of an executive assessment process conducted by a private global consulting firm. The sample is composed of those in upper management positions in corporations of varying sizes throughout the world. The corporations range from large, complex Fortune 50 companies with a global presence and household-name status to private equity-backed startup companies with few employees. The executives come from companies covering a broad spectrum of industries, from investment banks to steel companies to technology providers. Age and ethnicity-related data is not available for the executive group due to data privacy and anti-discrimination laws.

The subjects were required to take the 16PF for one of three reasons: 1) As part of a “Succession Management/Talent Review,” indicating a non-applicant status, 2) To aid in determining whether a current employee is appropriate for another role within the company (e.g. a promotion, or a move to another branch or division of the company), indicating an internal applicant status, or 3) To aid in the selection of an external candidate, again indicating an applicant status. For the purposes of this research, the external and internal candidate data will be considered in one group, as they both have a candidate status, and therefore also have a candidate mindset when completing the assessments.

Succession Management/Talent Review The assessment was given to some executives (N=1,223) as part of a talent review or succession management process,

during which a company wanted to get a better idea of the talent they already had in their ranks, as well as gain a greater understanding of the strengths and weaknesses of their executives in one or many areas of the business. These reviews are often used as part of a management change or for succession planning purposes. Those completing the 16PF as part of a Succession Management or Talent Review assessment were not under consideration for any specific role within the company, and were not part of any particular hiring or promotion process when they were assessed. Their results will only be used for the succession management/talent review project for which it was administered.

Job Candidate Job candidates (N= 449) were assessed as part of the selection process for a particular role either within the company at which the candidate currently works (internal candidates), or within companies by which the candidate had not previously been employed (external candidates). For internal candidates, the role may represent a traditional promotion, or may be a role within another business unit or region of a large organization. External candidates are traditionally new introductions to both the organization and the specific role.

The internal or external executive may be the only person in consideration for the role, or one of many from inside and/or outside the organization. All executives assessed are aware that they are being assessed in the context of their potential success in the specific role for which they are being considered, and that their candidacy for the role may be affected by their results.

Function Job function was recorded by the consulting firm. For the sake of parsimony, only the four most clearly defined functions were used. Those with a

“Finance” label may include CFOs, Treasurers, Controllers, Audit professionals, Financial Reporting executives, Investor Relations professionals, and others employed in a financial capacity. Those in the Marketing and Sales function include those with titles like Vice President of Marketing, Account Manager, Vice President of Sales, Sales Manager, etc. Operations executives may include those with titles like COO, Vice President of Operations, Plant Manager, etc. General Management executives typically serve in roles where a large part of the job description involves managing others or managing varied parts of the business. They may hold titles like Managing Director, CEO, or Division GM.

Industry The industry of the client company was determined by the consulting firm, typically using the primary SIC code to make determinations about primary industry.

Measure

The 16PF (Cattell, 1970) is a widely-used personality measure that consulting firms often use as part of an assessment battery. Initially published in 1949, the test is based on 16 facets of personality.

The 5th edition of the 16PF (Cattell, Cattell, & Cattell, 1993) consists of 16 scales determined using 10-15 items representing different areas of personality: Warmth (11 items), Emotional Stability (10 items), Dominance (10 items), Liveliness (10 items), Rule-Consciousness (11 items), Social Boldness (10 items), Sensitivity (11 items), Vigilance (10 items), Abstractedness (11 items), Privatness (10 items), Apprehension (10 items), Openness to Change (14 items), Self-Reliance (10 items), Perfectionism (10 items), Tension (10 items), and Reasoning (15 items). All scales are noncognitive with

the exception of Reasoning, which looks to measure basic cognitive ability.

Additionally, there is an “Impression Management” scale (IM, 12 items) designed to measure faking or response distortion levels, and specifically those within the “Impression Management” realm of Paulhus’s (1984) two-factor theory of Socially Desirably Responding.

Conn (1994) reports the mean test-retest reliability on the 16PF as .80, with a range of .69 (Reasoning) to .86 (Self-Reliance). The range of internal consistency is reported from .64 (Openness to Change) to .85 (Social Boldness) with a mean value of .74 (Conn, 1994).

CHAPTER 4

RESULTS

Hypothesis 1

Candidate scores on each of the 16 scales were compared to the scores of the incumbent group using a multivariate analysis of variance (MANOVA). 1,672 executives completed the 16PF and had a record of their candidate or incumbent status. The MANOVA yielded results significant at the $p < .001$ level for all three tests reported (Pillai's trace, Hotelling's T, Wilk's lambda), suggesting significant scale score differences exist between the candidate and incumbent groups on the 16 scales of the 16PF. Box's M was significant, which is a concern because of the large discrepancy in group sizes. However, because the determinant is larger for the larger group, we can ascertain that the tests were too conservative (Stevens, 2002). Because each of the tests was still significant despite the fact that each test was more conservative, the significance of Box's test is not problematic for this analysis.

In a post-hoc examination of each of the 16 16PF scales, 8 scales were determined to be statistically significant after Bonferroni corrections were applied to the univariate results. Univariate results for the scales are listed in Table 1, and means and standard deviations are found in Table 2. Reasoning (B), Emotional Stability (C), Social Boldness (H), Vigilance (L), Privatness (N), Openness to Change (Q1), Self-Reliance (Q2), and Tension (Q4) showed statistically significant differences between candidate and incumbent groups at an $\alpha = .003125$ level (corrected using Bonferroni). Can also be interpreted at a $p < .05$ level with a Bonferroni-adjusted critical F of 8.755).

Table 1

Univariate F-tests with (1,1672) D. F.

Bonferroni critical F = 8.755

*denotes that scale is statistically significant

Variable	Hypoth. SS	Error SS	Error MS	F
Warmth (A)	19.92777	4564.56686	2.73000	7.2995
Reasoning (B)	25.39449	3507.79846	2.09797	12.1043*
Emotional Stability (C)	38.61792	3179.48722	1.90161	20.3080*
Dominance (E)	6.44155	4469.86490	2.67336	2.4095
Liveliness (F)	21.21517	4449.56201	2.66122	7.9720
Rule Consciousness (G)	8.68909	4488.75715	2.68466	3.2366
Social Boldness (H)	106.88483	4813.32604	2.87878	37.1285*
Sensitivity (I)	6.54197	4127.10976	2.46837	2.6503
Vigilance (L)	54.29297	4315.36295	2.58096	21.0360*
Abstractedness (M)	1.03110	3393.62601	2.02968	.5080
Privateness (N)	65.99378	5342.65616	3.19537	20.6530*
Apprehension (O)	16.95970	3526.88977	2.10938	8.0401
Openness to Change (Q1)	72.86157	4810.21907	2.87693	25.3262*
Self-Reliance (Q2)	87.33818	5204.87269	3.11296	28.0563*
Perfectionism (Q3)	4.06772	5129.70588	3.06801	1.3259
Tension (Q4)	97.23119	3423.65351	2.04764	47.4845*

Table 2

Scale Means and Standard Deviations

*indicates scale was statistically significant in post-hoc univariate analysis

Scale	Candidate M	Cand Std. Dev.	Incumbent M	Incumb. Std. Dev.	Total Std. Dev.
A Warmth	5.90	1.469	5.66	1.712	1.654
B Reasoning*	7.43	1.328	7.15	1.489	1.452
C Emotional Stability*	7.13	1.200	6.79	1.439	1.387
E Dominance	6.95	1.490	6.81	1.683	1.634
F Liveliness	5.41	1.521	5.15	1.671	1.636
G Rule Consciousness	5.74	1.546	5.91	1.670	1.639
H Social Boldness*	6.90	1.626	6.33	1.723	1.716
I Sensitivity	4.48	1.428	4.34	1.620	1.572
L Vigilance*	3.69	1.563	4.09	1.621	1.615
M Abstractedness	4.63	1.415	4.68	1.429	1.425
N Privatness*	4.80	1.812	5.24	1.779	1.799
O Apprehension	4.31	1.393	4.54	1.473	1.455
Q1 Openness to Change*	7.30	1.544	6.82	1.544	1.749
Q2 Self-Reliance*	4.14	1.666	4.65	1.666	1.798
Q3 Perfectionism	5.66	1.665	5.55	1.782	1.751
Q4 Tension*	4.55	1.418	5.10	1.435	1.451

Correlations on the canonical variable created by the analysis were also interpreted. Emotional Stability (C), Social Boldness (H), and Openness to Change (Q1) all show a strong negative relationship with the canonical variable. Reasoning (B) shows a moderate negative relationship with the canonical variable. Vigilance (L), Privatness

(N), Self-Reliance (Q2), and Tension (Q4) all show strong positive relationships with the canonical variable. The same variables that were significant in the post-hoc univariate analysis represented the largest loadings on the canonical variable. Those scales on which the job candidates had a significantly higher mean score than incumbents had negative relationships with the canonical variable, and those scales on which incumbents scored significantly higher than candidates had positive relationships with the canonical variable. No other variables showed a notable relationship with the canonical variable.

In trying to understand patterns in the differences in results between the two groups, it is also important to examine the individual scales which showed the largest score differences, and determine what each of those scales is designed to measure. Factor B is a reasoning scale on the 16PF. In the results actor B showed the weakest relationship with the canonical variable, and had the lowest F-value of those scales that were significant. Job candidates obtained a higher score of the Reasoning scale than incumbents. Though the overwhelming majority of the candidates were employed elsewhere at the time they were being considered as candidates, they still did have more at stake in taking the 16PF. This may have led them to put more effort into completing the reasoning scale than incumbents. Because this was an ability scale and not a personality scale, it is likely the only scale upon which focus and effort can make a difference in results, as it is unlikely that in a relatively homogeneous sample, the job candidates would have significantly better reasoning ability than incumbents.

Candidates also received higher scores on Emotional Stability (C), which measures feelings about coping with everyday life, and has a strong correlation with the Impression Management scale (Russell & Karol, 2002). A low score on Emotional

Stability (C-) contributes to the anxiety global factor on the 16PF. The Social Boldness (H) scale, upon which the candidates also received a higher mean score, measures social boldness versus shyness, and contributes to the extraversion and independence global factors on the 16PF. This scale shows a strong relationship with self-esteem inventories. The final scale on which candidates obtained a significantly higher mean score than incumbents was the Openness to Change factor (Q1), which measures the extent to which an examinee is either open to innovation or experimentation or prefers the status quo. Openness to change contributes to the independence global factor and the receptiveness pole of the tough-mindedness global factor (Russell & Karol, 2002). Each of these factors also had a negative relationship with the canonical variable.

The incumbents recorded higher scores than the job candidates on four scales, each of which had a positive relationship with the canonical variable. Vigilance (L) is designed to measure the willingness to trust others rather than be vigilant about their motives or intentions (Russell & Karol, 2002). Vigilance contributes to both the anxiety and independence global factors. Privatness (N) measures the extent to which an examinee is forthright or private with others. Privatness loads negatively on the extraversion global factor on the 16PF, and is correlated with the Impression Management scale, with lower scales being more socially desirable (Russell & Karol, 2002). The level of desire to spend time alone versus being group-oriented is measured on the Self-Reliant scale (Q2). Self-Reliance has a negative correlation with the Impression Management scale, and loads on the extraversion global factor of the 16PF. Tension (Q4) has the largest correlation with the Impression Management scale ($r = -.53$; Conn & Rieke, 1994). According to the administration manual authors, "Social

desirability can affect Factor Q4 results. Since the items are fairly transparent, they can be influenced by response sets to present oneself favorably (Q4-)... (Russell & Karol, 2002, p.56)". Low scores indicate that an examinee is relaxed and patient, whereas high scorers may be restless and irritable. Tension also showed the largest F-statistic in the MANOVA analysis of the groups.

At least one of the scales significant at the univariate level loaded on all of the global factors of the 16PF (Extraversion, Tough-Mindedness, Independence, Anxiety) save the Self-Control global factor.

Hypothesis 2

A t-test examined group differences on the Impression Management scale of the 16PF. The t-test yielded a t-value of 1.789, which was not significant at the $p < .05$ level ($N=1,672$). Equal variances were assumed as Levene's test was not significant. Overall, the job candidate group had a higher mean score than the incumbent group, but the difference was not statistically significant.

Hypothesis 3

In the secondary analyses, the 16PF scores were broken down into four job functions that were categorized and recorded by the consulting firm. The functions of 1,057 executives were examined using a MANOVA. All three tests performed in the MANOVA (Pillai's Trace, Hotelling's T, Wilks Lambda) were statistically significant at the $p < .001$ level, supporting the hypothesis that differences between functional groups would exist. Box's M was significant, and the fact that the largest group (General Management) had the largest determinant, and the smallest group (Marketing & Sales) had the smallest determinant indicates that the MANOVA tests were too conservative

(Stevens, 2002). In an examination of the univariate analysis, Warmth (A), Social Boldness (H), and Openness to Change (Q1) were found to have significant differences between the functional groups after Bonferroni corrections were applied.

Hypothesis 4

The industry of the company at which the subject is a candidate or an incumbent was also examined. The data were categorized into the following types of companies: Industrial, Natural Resources, Consumer-Related, Technology, Financial Services, Healthcare, and Professional, Business, & Management Services. Differences between industry groups on the 16PF were examined using a MANOVA procedure. Because of the overpowering size of one group (the Industrial Companies group, $n=623$), these differences were examined using two different models: one which included all 7 industry groups, and one which excluded the exceptionally large group.

Model 1 A MANOVA was completed using 16PF and industry data from 1,711 executives in 7 groups. Box's M was significant at $p<.01$. Because the largest determinant in the model was that of the largest group, we can conclude that the MANOVA is being overly conservative (Stevens, 2002). Despite the conservative nature of the test, all three MANOVA tests (Pillai's trace, Hotelling's T, and Wilks lambda) were significant at the $p<.001$ level.

Model 2 Once again, Box's M was significant, this time at $p<.05$. The largest determinant in the model was again that of the largest group, we can conclude that the MANOVA is being overly conservative. All three MANOVA tests were again significant at the $p<.001$ level.

The models differed in the individual scales that were found to show significant univariate differences when a Bonferroni correction was applied. That discrepancy coupled with the issues raised by the discrepancies in group sizes and the size of the determinants of the various groups leads one to regard the results related to industry as questionable in their application. The data entry of this variable (completed by the consulting firm) may have suffered from some data quality and consistency issues, and thus these results will not be used to support the discussion.

CHAPTER 5

DISCUSSION

The job candidate and incumbent groups showed significant differences overall on the 16 scales of the 16PF personality measure. The statistical values supporting the difference were strong, showing a $p < .001$ significance level on the three most common MANOVA measures in spite of being subjected to particularly conservative testing conditions. Much of the research done on response distortion in personality measures contends that a difference between applicant and incumbent scores on a measure is an indication that response distortion may be occurring (Birkeland et al., 2001; Robie et al., 2001; Rossé et al., 1998). Whether those examinees are engaging in self-deception or impression management (or “other-deception”) as outlined by Paulhus (1984) is unclear, but given the prior inferences made from similar studies with similar results (e.g. Morgeson et al., 2007a; Rosse et al., 1998), one could argue that *some* kind of response distortion has occurred in this sample.

This assertion is further supported by the fact that a number of the scales that showed significant differences in a post-hoc univariate analysis have high correlations with the Impression Management, which is designed to measure socially desirable responding (SDR) and response distortion. Table 3 contains the correlations of each scale with the Impression Management scale in this study. Table 4 contains the original correlation of each of the 16PF scales with the Impression Management scale from the 16PF technical manual. Each table indicates which of 16PF scales were significant in the univariate analysis.

Table 3

16PF scale correlations with Impression Management (IM) scale

16 PF Scale	IM Correlation
Warmth (A)	.149*
Reasoning (B)	.032^
Emotional Stability (C)	.363* ^
Dominance (E)	-.014
Liveliness (F)	.032
Rule-Consciousness (G)	.217*
Social Boldness (H)	.208* ^
Sensitivity (I)	-.054
Vigilance (L)	-.354* ^
Abstractedness (M)	-.211*
Privateness (N)	-.181*^
Apprehension (O)	-.329*
Openness to Change (Q1)	.102* ^
Self-Reliance (Q2)	-.246* ^
Perfectionism (Q3)	.156*
Tension (Q4)	-.436* ^

*significant at $p < .01$ (N=1,799)

^significant univariate scale results in post hoc analysis

Table 4

16PF scale correlations with Impression Management (IM) scale
from (Conn & Rieke, 1994, p.64)

16 PF Scale	IM Correlation
Warmth (A)	.16*
Reasoning (B)	-.01 ^
Emotional Stability (C)	.50* ^
Dominance (E)	-.02
Liveliness (F)	-.09*
Rule-Consciousness (G)	.34*
Social Boldness (H)	.20* ^
Sensitivity (I)	-.03
Vigilance (L)	-.39* ^
Abstractedness (M)	-.36*
Privateness (N)	-.12*^
Apprehension (O)	-.39*
Openness to Change (Q1)	.06* ^
Self-Reliance (Q2)	-.21* ^
Perfectionism (Q3)	.17*
Tension (Q4)	-.53* ^

*significant at $p < .01$

^significant univariate scale results in post hoc analysis

Seven of the eight scales that were significant after a Bonferroni correction was applied have significant correlations with the IM scale ($p < .01$) where the scale scores of this study were correlated with one another, and six of the eight scales show significant correlations with the IM scale ($p < .01$) according to the norms of the technical manual (Conn & Rieke, 1994). The three scales with the strongest overall correlation with the Impression Management scale, both in this study and according to the norms (Tension, Emotional Stability, and Vigilance), were all significant in the univariate analysis, each producing F-statistic values greater than 20. In fact, the scale that yields the highest correlation with the Impression Management scale resulted in an F-statistic that was a full 10 units higher than the next closest statistic ($F = 47.48$). These results would support the notion that this difference between groups could exist as a result of response distortion or SDR. Given these scales differences, and the relationship of these scales with the Impression Management scale, reflected both in this study and given the norms that resulted from the original construction of the measure, we'd also expect to see a difference between the groups on the Impression Management scale.

On the contrary, although there was a difference in scores between the job candidate and incumbent groups on the 16PF, there was no significant difference between the groups on the Impression Management scale. According to the 16PF technical manual (Conn & Rieke, 1994), "To address the issue of socially desirable responding, the Impression Management (IM) scale was developed for the 16PF Fifth Edition...A high IM score indicates more socially desirable responding or exaggeration of socially desirable qualities... (p.60)" Furthermore, the 16PF administrator's manual (Russell & Karol, 2002) states that,

“IM is essentially a social-desirability scale... Thus high scores can reflect impression management (presenting oneself to others as tending to behave in desirable ways) or they can reflect an examinee’s self-image as a person who behaves in desirable ways. In both cases, the possibility exists that the socially desirable responses might be more positive than the examinee’s actual behavior (i.e. a form of response distortion that may be conscious or unconscious)...(p.22)”

With that in mind, one would think that in a large sample, the groups would show a significant difference on the Impression Management scale if they showed a difference on the 16PF scales, and more importantly a difference on *half* of the 16PF scales on a univariate level, and even further on specific scales with a strong correlation with the Impression Management scale. Yet no significant difference is seen. Given that the stated purpose of the Impression Management scale is “to address the issue of socially desirable responding (Conn & Rieke, 1994, p.60),” and its description contains verbiage about its use as measurement tool for a “form of response distortion (Russell & Karol, 2002; p. 22),” if the differences between groups are in fact due to response distortion or social desirable responding, the fact that the scale may not be capturing that very phenomenon across a large sample is disconcerting,.

That said, these results cannot speak directly to the validity of the IM scale of the 16PF. However, it can be said that if the IM scale of the 16PF was *not* valid, we would see no significant difference between the two groups on the scale when these types of differences exist on the other 16 scales of the 16PF. Even having the ability to support that statement should lead those creating these personality measures to reexamine social

desirability scales, and determine whether or not they are successful in measuring what they are designed to measure.

The fact that an analysis of scale differences on the 16PF among 4 different functions was significant as well was also of interest. Though Hypothesis 3 did predict a difference between functional groups, the difference that appeared does not present a convincing case of response bias. Of the three univariate scales that were determined to be significant, only one (Social Boldness) has a correlation with the Impression Management scale of .20 or greater in either this study or according to the norms ($r = .208$; $r = .20$; Conn & Rieke, 1994). Given that different personalities may be drawn to different professions (Miller, 1962), it is not surprising that some differences existed, and that the Impression Management scale itself was not significant in this case (with an extremely low F of 0.357), as real personality differences have been found to exist between occupational groups (Daniels & Hunter, 1949). But when all occupational areas are taken together, their differences should complement one another, creating a relatively normal distribution. However, in the analysis of the candidate and incumbent groups, we see many more scales that are significant, that different scales are significant (Warmth carried a significant difference in the analysis of job function, but showed no significant difference between the candidate and incumbent groups), and that majority of those significant scales have a strong correlation with the IM scale.

The effectiveness of Impression Management and Socially Desirability scales in detecting response distortion and socially desirable responding (SDR) has been discussed at great length since the creation of the scales (Edwards, 1957; Holden, 2008). There are

those that say that the scales do exactly what they are designed to do (Cattell & Cattell, 1995), but the results of the present study would not support that notion if the differences between groups are, in fact, a result of response distortion or SDR.

On the other hand, some have argued that these scales may measure constructs other than response distortion. Ones, Viswesvaran, & Reiss (1996) conducted a meta-analysis of social desirability literature, and ultimately came to the conclusion that socially desirability scales were not a measure of SDR insofar as it relates to response distortion. Instead, they claim that “social desirability scales reflect individual differences in personality variables (p.669). In that sense, their definition of social desirability differs from much of the literature in that it contends that social desirability is not a measure of response bias or distortion.

That definition-changing assertion leads me to a few relevant questions: What is the point of including a scale in a personality measure that designed to do one thing, but may actually do something else? If, in fact, the scale is a better measure of another construct, why defend its continued use in its intended and stated capacity? And would the debate exist if instead of involving Impression Management scales, it turned out that an Emotional Stability scale was better at measuring Vigilance than Emotional Stability? Much of the debate about these social desirability scales seems to be influenced by the overarching debate about the effectiveness of self-report personality measurements, and whether or not their validity is affected by response distortion.

If those creating personality measures take the definition of a social desirability scale down the same route as Ones et al. (1996), and wish to create a scale that measures the extent to which an examinee attempts to make a good impression or please others,

that would change the intended purpose of social desirability scales. However, this change would make the social desirability scale a facet of the personality portion of the measure, rather than a tool used to determine whether or not applicants are distorting their responses. If that change is made, the stated purposes of these new scales would not be able to claim that scale scores can be used as a method by which to measure response distortion.

Elliott (1981) proposed changing the way in which the scales are interpreted. Instead of using the scales as a measure of response bias or response distortion, Elliott suggested they should serve as a measure of conscientiousness, and can assist in determining whether or not the test-taker wants to make a good impression. Extraordinarily high scores may then still be brought to the attention of test administrators, alerting them to the fact that particular test-takers may be trying too hard to make the proper impression. But, marketing the scale as a cut-and-dry measure of response distortion levels would be ignoring approximately 15 years of research that has shown that the scales may not measure response distortion and response distortion alone, and is instead measuring a number of constructs, one of which may be response distortion or “faking.”

Some may regard this as an argument about semantics, but in the realm of personality research as it relates to Industrial and Organizational psychology, a correction or clarification of the semantics used would be a step in the right direction. Though the literature seems to have exhausted the debate on SDR’s effect on personality measures, no substantial movement toward change has occurred as a result of that debate. Because the debate seems to be reigniting in the wake of recent publications (e.g. Morgeson et al.,

2007), steps should be taken to try to make self-report personality measures better. Acknowledging social desirability scales and impression management scales for what they really are would be one small step in the right direction, and perhaps encourage additional improvements. If personality measures are improved to remove some of the gray area which has sparked so much debate over the last 50 years, we would ultimately have better measures. One would hope these improved measures would lead to better results, which would leave less room for argument over the validity and meaning of the results obtained.

Limitations

There is no way to determine with certainty that the differences in the groups are a direct reflection of response distortion or socially desirable responding. Though prior research has often compared applicant and incumbent samples to show evidence of response distortion, the results are still not conclusive, as it is possible that other factors are influencing the results.

Firstly, it is possible that real differences exist between the candidate and incumbent groups in this sample. The sample is not random, and although the vast majority of those in the candidate group are currently employed, it may be possible that true personality differences exist between those currently in roles, and those who are considering leaving their current role to pursue a new position. The incumbents may also be showing frame-of-reference effects when compared to the candidates, as they are able to respond to items with the knowledge of what their current position entails, and how they have reacted to certain situations while within that role in the past, whereas

candidates are forced to use the knowledge they have to try to determine how they would act in a role they haven't yet held.

Alternatively, it is possible that members of the incumbent group could have also viewed their assessment as a high-stakes situation, and in turn also engaged in socially desirable responding or response distortion. This group could have potentially seen the assessment as something that could put their jobs at risk if their responses were not deemed to be acceptable, and in turn, these results could stem from the fact that both groups were engaging in response distortion, which would explain the fact that there was no difference between the group on the Impression Management scale, but would not necessarily explain the univariate differences on scales that are correlated with the Impression Management scale.

Future Directions

A logical direction for an expansion of this study would examine the amount of variance in personality scale scores accounted for by function or industry as compared to the amount accounted for by candidate/incumbent status. Attempting to determine which factor has a greater contribution to the variance in the personality scale scores could shed light on the strength of SDR, determining whether factors like function and industry carry more weight in the scale scores of personality measures than whether or not the examinee is under direct consideration for a role or not when completing the measure.

REFERENCES

- Azjen, I. (1991). Theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Barrick, M., & Mount, M. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81(3), 261-272.
- Birkeland, S., Manson, T., Kisamore, J., Brannick, M., & Smith, M. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317-335.
- Bradley, K., & Hauenstein, N. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science*, 48(3), 313-335.
- Braun, J., & Lafaro, D. (1968). Fakability of the sixteen personality factor questionnaire, Form C. *Journal of Psychology: Interdisciplinary and Applied*, 68(1), 3-7.
- Cattell, R., & Cattell, H. (1995). Personality structure and the new fifth edition of the 16PF. *Educational and Psychological Measurement*, 55(6), 926-937.
- Cattell, R., Eber, H., & Tatsuoka, M. (1970). *Handbook for the 16PF Questionnaire*. Champaign, IL: IPAT.
- Christiansen, N., Goffin, R., Johnston, N., & Rothstein, M. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47(4), 847-860.
- Conn, S., & Rieke, M. L. (Eds.). (1994). *The 16PF fifth edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing.

- Damarin, F., & Messick, S. (1965). *Response styles as personality variables: A theoretical integration* (ETS RB 65-10). Princeton, NJ: Educational Testing Service.
- Daniels, E., & Hunter, W. (1949). MMPI personality patterns for various occupations. *Journal of Applied Psychology, 33*(6), 559-565.
- Dilchert, S., Ones, D., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science, 48*(3), 209-225.
- Edwards, A. (1957a). Social desirability and probability of endorsement of items in the interpersonal check list. *The Journal of Abnormal and Social Psychology, 55*(3), 394-396.
- Edwards, A. (1957b). *The social desirability variable in personality assessment and research*. Fort Worth, TX: Dryden Press.
- Ellingson, J., Smith, D.B., & Sackett, P. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*(1), 122-133.
- Elliott, A.G. (1981). Some implications of lie scale scores in real-life selection. *Journal of Occupational Psychology, 54*(1), 9-16.
- Guion, R. M., & Cottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135-164.
- Heggestad, E., Morrison, M., Reeve, C., & McCloy, R. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*(1), 9-24.

- Hogan, J., Barrett, P., & Hogan, R (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92*(5), 1270-1285.
- Holden, R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences, 44*(1), 311-321.
- Hough, L. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*(2), 209-244.
- Hough, L., Eaton, N., Dunnette, M., Kamp, J., & McCloy, R. (1990). Criterion-related validity of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
- Krug, S. (1978). Further evidence on 16 PF distortion scales. *Journal of Personality Assessment, 42*(5), 513-518.
- Leary, M., & Kowalski, R. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*(1), 34-47.
- Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*(3), 226-246.
- McFarland, L., & Ryan, A (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812-821
- McFarland, L., & Ryan, A. (2001). Toward an integrated model of applicant faking behavior. Paper presented at the 2001 conference of the Society for Industrial and Organizational Psychology. San Diego, CA.
- Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: The role of over-claiming, self-esteem, and emotional intelligence. *Psychology Science, 48*(3), 336-356.

- Messick, S., & Jackson, D. (1961). Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin*, *58*, 299-304.
- Miller, S. (1962). Relationship of personality to occupation, setting, and function. *Journal of Counseling Psychology*, *9*(2), 115-121.
- Morgeson, F., Campion, M. & Dipboye, R., Hollenbeck, J., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*(3), 683-729.
- Morgeson, F., Campion, M. & Dipboye, R., Hollenbeck, J., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, *60*(4), 1029-1049.
- Mueller-Hanson, R., Heggestad, E., & Thornton, G. (2006). Individual differences in impression management: An exploration of the psychological processes underlying faking. *Psychology Science*, *48*(3), 288-312.
- Ones, D., Dilchert, S., Viswesvaran, C., & Judge, T. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, *60*, 995-1027.
- Ones, D., Viswesvaran, C, & Reiss, A. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.
- Paulhus, D. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.
- Paulhus, D. (2002). Socially desirable responding: The evolution of a construct. In H.L

- Braun, D.N. Jackson, & D.E. Wiley. *The role of constructs in psychological and educational measurement*. Mahwah, NJ: Erlbaum.
- Robie, C., Zickar, M., & Schmit, M. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance, 14*(2), 187-207.
- Rossé, J., Stecher, M., Miller, J., & Levin, R (1998) The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.
- Russell, M., & Karol, D. (2002). *16PF fifth edition administrator's manual*. Champaign, IL: Institute for Personality and Ability Testing.
- Sackeim, H., & Gur, R. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology, 47*(1), 213-215.
- Schmit, M., Ryan, A., Stierwalt, S., & Powell, A. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607-620.
- SHL Group (1999). *OPQ32 Manual and User's Guide*. Thames Ditton, UK: SHL Group.
- Smith, D., Hanges, P., & Dickson, M. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology, 86*, 304-315.
- Snell, A., Sydell, E., & Lueke, S. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review, 9*, 219-242.
- Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.)

Mahwah, NJ: Erlbaum.

Tett, R.; Christiansen, N.;(2007) Personality tests at the crossroads: A response to

Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007).

Personnel Psychology, 60(4), 967-993.

Wiggins, J.S. (1964). Convergences among stylistic response measures from objective

personality tests. *Educational and Psychological Measurement*, 24, 551-562.