Illinois State University

ISU ReD: Research and eData

Theses and Dissertations

1-13-2015

Personality Test Faking: Detection and Selection Rates

David J. Wolfe

Illinois State University, djwolfe@ilstu.edu

Follow this and additional works at: <https://ir.library.illinoisstate.edu/etd>

Part of the [Psychology Commons](#)

Recommended Citation

Wolfe, David J., "Personality Test Faking: Detection and Selection Rates" (2015). *Theses and Dissertations*. 298.

<https://ir.library.illinoisstate.edu/etd/298>

This Thesis is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISUREd@ilstu.edu.

PERSONALITY TEST FAKING:
DETECTION AND
SELECTION
RATES

David J. Wolfe

200 Pages

May 2015

This study examined the utility of Kuncel & Borneman's (2007) novel approach to faking detection using unusual item responses, after having addressed several limitations of their previous study. Their approach was applied to a group of Romanian professionals that took a personality test (the NEO-PI-R) on two occasions 12-24 months apart. This within-subjects design using real job applicants allowed for evaluation of faking at real-world individual levels, as well as offered the ability to analyze Kuncel and Borneman's (2007) proposed technique with a prevalent selection tool that uses a more conventional five-option response set. Following the theory proposed by Griffith, Chmielowski, and Yoshita (2007), confidence intervals were calculated and used to determine the faking behavior of the individuals in the study. Results suggested that the Kuncel and Borneman (2007) method of faking detection was amenable to a real-world application context, and that a quantitative method of unusual item recoding was superior to the previously used qualitative approach.

PERSONALITY TEST FAKING:
DETECTION AND
SELECTION
RATES

DAVID J. WOLFE

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Psychology

ILLINOIS STATE UNIVERSITY

2015

© 2015 David J. Wolfe

PERSONALITY TEST FAKING:
DETECTION AND
SELECTION
RATES

DAVID J. WOLFE

COMMITTEE MEMBERS:

Dan Ispas, Chair

Alexandra Ilie, Co-Chair

Suejung Han, Reader

ACKNOWLEDGMENTS

The writer would like to express deep appreciation to my committee chair, Dr. Dan Ispas, whose experience and exigent nature resulted in the continual development of my quest toward both research and scholarship. Without his guidance, highest of standards, and attention to detail, this thesis would not have been realized.

I would also like to thank my co-chair, Dr. Alexandra Ilie, and reader, Dr. Suejung Han, whose dedication to the field of Psychology was evidenced in their necessarily tireless and insightful efforts toward improving the quality of this work, through reading, critiquing, and suggesting directions toward which to aim the nature of my study.

Additionally, gratitude toward Dr. John Binning, who spent many hours introducing me to this area of Psychology, and developing within me the interest and desire to advance such research with a project of this magnitude. I would also like to thank Dr. Eros DeSouza, who took a personal interest in my advancing education from the earliest possible time, and whose initial and timely encouragement ensured that I reached my goals on schedule.

Thanks to Ashley McCarthy, Sam Hayes, and Ryan Tuggle for assisting with item ratings, making an analysis of interrater agreement possible. Thanks to Hannah Archos for help with discussing ideas, reviewing grammar, and aiding in proofreading multiple

drafts of this project. Finally, thanks to my family for always encouraging me in the completion of this goal, and for continuously making adjustments in their own lives to accommodate me whenever I asked.

D.J.W.

CONTENTS

ACKNOWLEDGMENTS	i
CONTENTS	iii
TABLES	vi
FIGURES	ix
CHAPTER	
I. THE PROBLEM	1
Statement of the Problem	1
II. REVIEW OF RELATED LITERATURE	3
The Predictive Power of Personality Assessment	3
Predictive Validity	3
Incremental Validity	10
Effects on Adverse Impact	11
Criticism Regarding the Use of Personality Measures in Selection Contexts	12
Faking and Personality Assessment	16
Directed-Faking in Laboratory Studies	18
Applicant Research in High-Stakes Contexts	20
Faking: Insignificant Problem of Legitimate Concern?	23
Is Faking Socially Adaptive?	23
What is the Prevalence of Faking?	25
Does Faking Affect the Predictive Validity of Personality Measures?	28
The Impact of Faking on Selection Rates and Hiring Decisions	30
Faking and Select-In Hiring Decisions	31

Faking and Select-Out Hiring Decisions	35
Previous Approaches Used to Address Concerns Regarding Potential Faking	37
Methods That Attempt to Control or Eliminate the Problem	38
Methods That Attempt to Detect the Problem	43
The Kuncel and Borneman (2007) Unusual Item Response Technique	52
III. SUMMARY AND RESEARCH QUESTIONS	60
Summary	60
Research Questions	62
IV. METHOD	63
Participants	63
Measures	63
Procedure	70
Research Questions 1A and 1B	70
Exploratory Inter-rater Agreement	73
V. RESULTS	88
Descriptive Statistics	88
Reliabilities	88
Correlations Between Research Factor Scores and Faking Indicators	89
Factor Score Changes (Between Applicant and Research Conditions)	90
Research Question 2	92
1 SD Categorization Method	93
½ SD Categorization Method	94
Research Question 3	96
1 SD Categorization Method	96
½ SD Categorization Method	98
Research Question 4	102

Conscientiousness/ 1 SD	102
Conscientiousness/ ½ SD	105
Neuroticism/ 1 SD	107
Neuroticism/ ½ SD	110
Extraversion/ 1 SD	113
Extraversion/ ½ SD	116
Research Question 5	117
Conscientiousness/ 1 SD	117
Conscientiousness/ ½ SD	120
Neuroticism/ 1 SD	122
Neuroticism/ ½ SD	124
Extraversion/ 1 SD	127
Extraversion/ ½ SD	129
Exploratory Curvilinear Analysis	131
1 SD Categorization Method	131
½ SD Categorization Method	134
VI. DISCUSSION	138
Summary of Findings	138
Strengths	147
Limitations	149
Implications for Practice	151
Implications for Research and Theory	153
VII. CONCLUSION	157
REFERENCES	158
APPENDIX A: Decomposition of True Faking Categorization Methods	171
APPENDIX B: Figures Depicting Comparisons of the Respective Methods	176

TABLES

Table	Page
1. Sample Recoding Scheme for One Item (Kuncel & Borneman, 2007)	57
2. Descriptive Statistics for the 42 Unusual Items, with Contrasts from the Research Condition to the Applicant Condition	77
3. Sample Recoding Scheme for Item 21 Representing the Impulsiveness Facet of Neuroticism	81
4. Preliminary Findings Regarding Applicability of Various Methods for Categorizing True Fakers	83
5. Correlations Between NEO-PI-R Factor Results from the Research Condition and the Respective Faking Indicator Scores (Quantitative and Qualitative)	90
6. Paired-Samples t-Test Results for Differences Between Conditions for Each of the Five Personality Factors, Along with Means and Standard Deviations from the Respective Conditions	92
7. 1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Quantitative Faking Indicator	94
8. ½ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Quantitative Faking Indicator	96
9. 1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Kuncel and Borneman (2007) Qualitative Faking Indicator	98
10. ½ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Kuncel and Borneman (2007) Qualitative Faking Indicator	100
11. Differences in 1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Between my Quantitative Faking Indicator and the Kuncel and Borneman (2007) Qualitative Indicator	101

12.	Differences in ½ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Between my Quantitative Faking Indicator and the Kuncel and Borneman (2007) Qualitative Indicator	101
13.	Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Conscientiousness	105
14.	Impact on Select-In Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Conscientiousness	107
15.	Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Neuroticism	110
16.	Impact on Select-In Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Neuroticism	113
17.	Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Extraversion	115
18.	Impact on Select-In Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Extraversion	116
19.	Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Conscientiousness	119
20.	Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Conscientiousness	122
21.	Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Neuroticism	124
22.	Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Neuroticism	126

23.	Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Extraversion	129
24.	Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Extraversion	130
25.	Impact on Curvilinear Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and All Three Predictors	134
26.	Impact on Curvilinear Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and All Three Predictors	136

FIGURES

Figure	Page
1. A Typical Item's Response Distributions from Honest (a) and Faking (b) Conditions for the Test Item Careful	54
2. An Unusual Item's Response Distributions from Honest (a) and Faking (b) Conditions for the Test Item Imperturbable	55
3. An Unusual Item's Response Distributions from Research (a) and Applicant (b) Conditions for Item 123 Representing the Fantasy Facet of Openness	72
4. An Unusual Item's Response Distributions from Research (a) and Applicant (b) Conditions for Item 21 Representing the Impulsiveness Facet of Neuroticism	73

CHAPTER I

THE PROBLEM

Statement of the Problem

Despite some early opposition (Ghiselli & Barthol, 1953; Guion & Gottier, 1965), personality assessment has become a vital component in the practice of Industrial/Organizational (I/O) Psychology (Rothstein & Goffin, 2006). However, there is a question as to whether self-report measures of personality are also susceptible to faking behaviors (Ziegler, MacCann & Roberts, 2011). Although an array of methods intended to control for, reduce, or eliminate the possibility of faking have been investigated, there is no widely accepted solution to this potential problem to date (Kuncel & Borneman, 2007; Reeder & Ryan, 2011). This paper will attempt to provide a review of the situation, and address limitations of Kuncel & Borneman's (2007) recent method for detecting faking on personality measures in selection contexts.

I will begin with an historical review of the use of personality assessments for selection purposes in I/O Psychology. I will then discuss the divergent perspectives regarding the susceptibility of such measures to faking. Next, I will review research that has examined the impact of faking on selection rates and hiring decisions. I will follow that review with a discussion of various methods researchers have proposed for controlling or detecting faking behaviors. After that, a thorough elaboration of one study

that used a novel method will be offered, including a discussion of some of its notable limitations. Finally, I will elaborate on the nature of the current study, which will address these limitations in an attempt to determine the practical utility of this novel approach to faking detection.

CHAPTER II

REVIEW OF RELATED LITERATURE

The Predictive Power of Personality Assessment

Predictive Validity

According to Schmidt and Hunter (1998), from a practical perspective the most important part of personnel assessment is its predictive ability. It is often reported in the literature (and commonly accepted amongst professionals) that measures of general mental ability (GMA) or cognitive ability offer the best or most valid prediction of job performance (Hunter & Hunter, 1984; Jensen, 1998; Ree & Earles, 1992; Schmidt & Hunter, 1998). However, hiring based on cognitive ability alone has been shown to have adverse impact on minority groups (Hunter & Hunter, 1984). As it is an important concern of many organizations (one with legal ramifications in the United States), there is extensive research regarding effective hiring strategies that also reduce such adverse impact (Newman, & Lyon, 2009; Ployhart & Holtz, 2008; Pulakos & Schmitt, 1996; Ryan, Ployhart, & Friedel, 1998). A commonly used method with which to accomplish this is to add a non-cognitive measure (most often some aspect of personality) as a predictor (Avis, Kudisch, & Fortunato, 2002; Potosky, Bobko, & Roth, 2005; Schmidt & Hunter, 1998).

Although today it is quite common to witness the use of personality measures in selection contexts, this was not always the case (Rothstein & Goffin, 2006). Early reports, such as Ghiselli and Barthol's (1953) review of personality and selection (that found mixed results and counterintuitive findings), may have served to suppress their use. With their comprehensive review of personality assessments, Guion and Gottier (1965) may have provided most of the impetus in pushing the field of I/O Psychology away from personality measurement when they reported that none of the personality measures that were examined demonstrated usefulness in selection contexts. While admitting evidence of the need for organizations to predict personality-related behaviors deemed important to the job, they concluded that there was a lack of evidence suggesting that the use of such tests could be recommended as practical tools for selection purposes (Guion & Gottier, 1965). The authors further concluded that the only situation in which the use of personality measures as selection tools was acceptable was after extensive research into their use in a specific situation for that specific purpose (Guion & Gottier, 1965). Although the authors also acknowledged that the same could be said of any predictor, they suggested that the problem was even more severe with measures of personality (Guion & Gottier, 1965).

This report led to what Hough and Oswald (2008) regarded as a twenty year lull in research regarding personality for I/O Psychology. One influence toward the return of personality measurement in work psychology was the United States Army's seven-year undertaking in the 1980's known as Project A, which was designed to develop a selection and classification system for the entirety of the organization (Campbell, 1990; Hough &

Oswald, 2008). Its construct-based approach resulted in a concurrent validation study that found that personality measures, when linked to relevant constructs and correlated with job analysis scores, do offer predictive ability (Hough & Oswald, 2008; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990).

Contemporaneous advances in personality theory, such as the agreement on five robust factors of personality (Neuroticism or Emotional Stability, Extraversion, Openness, Agreeableness, and Conscientiousness) and Costa and McCrae's 1985 development of a personality inventory designed to measure them (the NEO-PI), also began to emerge (Costa & McCrae, 1992; Digman, 1990; Murphy, 2005). The NEO-PI measures the five factors of personality using six facets represented by 30 scales comprised of 8 items each (Costa & McCrae, 1992). Such progress influenced others to increase their own efforts toward establishing the utility of personality in work contexts, as evidenced by Hogan and Hogan's (1989) development of the Employee Reliability scale. In fact, much of the subsequent research has highlighted the importance of these five factors in occupational settings, with Conscientiousness and Emotional Stability the most prominently cited (Barrick & Mount, 1991; Hogan & Holland, 2003; Salgado, 1997; Tett, Jackson, & Rothstein, 1991). For this reason, the current research will focus on these two factors in its analyses.

Such research may be considered as beginning with early meta-analytic work regarding personality and occupational outcomes. While citing the recent convergence of personality psychologists on the five-factor model (FFM) of personality, Barrick and Mount (1991) published their landmark meta-analysis. Reviewing over 30 years of

research, the authors focused on articles that examined the five factors as they related to multiple performance measures in a variety of occupations (Barrick & Mount, 1991). Because they deemed the characteristics that comprise it as important to accomplishing tasks in all jobs, the authors predicted that Conscientiousness would represent the most valid predictor of job performance (Barrick & Mount, 1991). According to Barrick and Mount (1991), these characteristics include: dependability; being careful, thorough, responsible, organized, and planful; and variables of personal volition such as hardworking and perseverance.

The results supported their hypothesis, indicating that Conscientiousness was a valid predictor across all occupational categories and all criterion types included in their meta-analysis (Barrick & Mount, 1991). This was consistent with earlier results of Project A and Hogan and Hogan's (1989) Employee Reliability scale which respectively found that aspects of Conscientiousness, and a measure that was heavily based on such aspects, were valid predictors of job performance (McHenry et al., 1990). Tett et al.'s (1991) meta-analysis (which used a slightly different method and reported similar results with even higher validities for all five factors) provided further support for continuing the investigation into the role of personality in performance prediction.

Since that time, the field has witnessed an explosion of publications regarding the FFM and its predictive ability in work contexts, often with an emphasis on Conscientiousness (Avis et al., 2002; Behling, 1998; Costa, 1996; Dudley, Orvis, Lebiecki, & Cortina, 2006; Hogan, & Holland, 2003; Hurtz & Donovan, 2000; Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007a; Morgeson,

Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007b; Salgado, 1997). Extending the research of Barrick and Mount (1991) and Tett et al. (1991), Salgado (1997) conducted his own meta-analysis regarding the FFM and work within the European community. He reported that Conscientiousness and Emotional Stability (or reverse-scored Neuroticism, which can be defined as being: anxious, depressed, angry, embarrassed, emotional, worried, and/or insecure) were valid predictors across multiple criteria and occupations (Barrick & Mount, 1991; Costa & McCrae, 1992; Hills & Argyle, 2001; Salgado, 1997; Ziegler et al., 2011).

Hurtz and Donovan's (2000) meta-analysis sought to extend this research by including only studies that used scales explicitly designed to represent the FFM. They reported that Conscientiousness evidenced the highest predictive ability, with Emotional Stability generally the next highest (Hurtz & Donovan, 2000). Barrick, Mount and Judge (2001) upheld the notion that Conscientiousness is a valid predictor of all criterion types (for work performance) across all occupations in their summary of the recent explosion of meta-analytic publications regarding the FFM and work performance. The authors further concluded that Emotional Stability was a valid predictor of overall work performance and teamwork across jobs, although this predictor was not as consistent nor was it as strong as was Conscientiousness (Barrick et al., 2001).

In another meta-analysis, Hogan and Holland (2003) found that all dimensions of the FFM predicted theory-relevant criteria, increasingly so as the criteria became more specific. Previous studies had reported similar findings, such as evidence that Extraversion is a valid predictor for occupations involving social interaction (Barrick &

Mount, 1991; Salgado, 1997). In addition, the authors made special note of what they deemed the previously under-realized, high predictive ability of Emotional Stability under such theory driven examination (Hogan & Holland, 2003). Further, the results of Dudley et al.'s (2006) meta-analytic work regarding Conscientiousness at the facet level suggested that Dependability might be the most important facet driving Conscientiousness across criteria and occupations. This echoes previous results from Project A and Hogan and Hogan (1989) in which the dependability facet of Conscientiousness factored prominently (McHenry et al., 1990).

A recent meta-analysis from Judge, Rodell, Klinger, Simon, and Crawford (2013) further explored the role of bandwidth in the links between personality and job performance. They examined the effects of prediction using intermediary traits (between the higher-order factors and lower-order facets) within the factors of the FFM and found that moving from broad level traits to the more narrow traits often resulted in changes in predictive ability (Judge et al., 2013). Overall, the authors found that Conscientiousness generally evidenced the highest correlations with work outcomes of overall, task, and contextual performance (Judge et al., 2013). This trait as a single aggregate predicted these criteria better than its two intermediary traits (industriousness and orderliness), which themselves often evidenced higher correlations (as well as less variance than that evidenced among the individual facets) than all except the highest correlated facet (Judge et al., 2013). Neuroticism was mostly negatively correlated with work outcomes, with its two intermediary traits (volatility and withdrawal) varying to a greater degree than those of Conscientiousness (Judge et al., 2013). Further, volatility actually evidenced higher

correlations with all three criteria than did Neuroticism as a single aggregate (Judge et al., 2013). In addition, volatility often outperformed all but the highest correlated individual facets, while withdrawal often performed at a level at or below the single aggregate and most of the facets alone (Judge et al., 2013). In summary, the data from this study for these two factors suggest that sometimes narrow can be too narrow, while slightly more narrow may prove valuable depending on the linkage being examined (Judge et al., 2013).

What remains clear is that further investigation into the nature of such linkages has the potential to clarify questions regarding both theoretical relationships between the factors and facets (and perhaps even intermediary traits) and various levels of criteria, as well as what may be the appropriate bandwidth for various predictive contexts.

Relatedly, Li, Barrick, Zimmerman, and Chiaburu (2014) argued that broad factors must be matched with equally broad criteria to maximize their effectiveness (Li et al., 2014).

While they concede that the prediction of narrow outcomes often requires narrower predictors, they stressed that broad predictors such as the factors that comprise the FFM predict generalized work behaviors better than has been acknowledged (Li et al., 2013).

Relatedly, Dudley et al.'s (2006) meta-analysis reported that narrow facets were better predictors than global Conscientiousness for all criteria except overall job performance.

The general rule may simply be that the bandwidth of the predictor is often best when matched to that of the criterion.

Incremental Validity

Continuing the discussion regarding the utility of personality measurement in selection contexts, not only have personality measures evidenced predictive validity, they have also actually been shown to add to the predictive ability of selection systems above and beyond using cognitive ability alone. For instance, Day and Silverman (1989) found that several job-relevant aspects of personality predicted ratings of job performance above cognitive ability alone. Judge, Higgins, Thoresen, and Barrick (1999) later reported that the five factors of the FFM offered incremental validity above cognitive ability regarding job satisfaction and career success. Avis et al. (2002) reported incremental validity over cognitive ability alone regarding multiple performance criteria when using Conscientiousness as a predictor.

Furthermore, in an article detailing three meta-analyses regarding cognitive ability and personality, Salgado (1998) reported incremental validity of Conscientiousness and Emotional Stability over cognitive ability alone across civil and military occupations. Schmidt and Hunter (1998) reported an 18% increase in validity when adding Conscientiousness to GMA, while stating that such incremental validity translates into increases in “practical value” (p. 266). The authors also reported that integrity tests, (which they stated measure mostly Conscientiousness, along with Agreeableness and Emotional Stability) provide a 27% increase in validity above GMA alone (Schmidt & Hunter, 1998). As Hough and Oswald (2008) state in summary, not only do personality variables show incremental validity, they thereby increase the accuracy of predictions of performance and other job-related criteria.

Effects on Adverse Impact

In their summary of research regarding adverse impact and mean subgroup differences, Hough, Oswald, and Ployhart (2001) noted very little difference between ethnic/cultural and age groups at the factor level of the FFM, while cautioning that sample sizes for American Indians and Asian Americans were too small to be conclusive. Notwithstanding, they also reported consistent differences between such groups at the facet levels of Conscientiousness and Extraversion (Hough et al., 2001). Additionally, they reported that noteworthy differences between genders at both the factor (women score higher than men on Agreeableness) and facet (under Conscientiousness, women score higher than men on Dependability but lower on Achievement) levels appear to exist (Hough et al., 2001). The authors concluded that practically meaningful differences between groups are moderated by the specific construct (factor or facet) in question, and that appropriate job analysis and attention to job requirements is necessary when employing measures of personality (Hough et al., 2001). While reaffirming the importance of job analysis information to identify important abilities and characteristics, Hough and Oswald's (2008) review of personality use in I/O Psychology asserted that selection batteries comprised of only personality measures "typically (but not always) satisfy the four-fifths rule" (p. 285).

Indeed, several studies that did not support the theory that adding non-cognitive predictors to cognitive tests reduces adverse impact have been published. Ryan et al. (1998) reported that the use of personality did not ameliorate adverse impact effects for two samples of applicants for police and firefighting work. They suggested that applicant

pool characteristics (such as standard deviation and mean differences between groups) and the personality test that was used might be more important determinants regarding the effects of adverse impact (Ryan et al., 1998). As mentioned above, Avis et al. (2002) found that Conscientiousness offered incremental validity over cognitive ability alone for multiple performance criteria, although they also reported that it actually worsened adverse impact. In a similar study, Potosky et al. (2005) found that adding Conscientiousness resulted in only a slight increase in validity over cognitive ability alone while producing negligible differences in expected adverse impact. It may be that more investigation into the moderating and mediating effects of attentive job analysis and care in choosing specific constructs will increase the efficacy of such instruments in reducing adverse impact. However, along with myriad other concerns, such findings have contributed to the impetus causing some to question the current use of personality measurement in I/O Psychology.

Criticism Regarding the Use of Personality Measures in Selection Contexts

Murphy (2005) offered several additional potential problems with the use of personality inventories for selection purposes. While pointing to what he deemed as the relatively low levels of criterion-related validity associated with tests of personality, he suggested that: the theories linking personality constructs with job performance are often weak; little is known about how best to match personality attributes with varying occupations; and that the measures used to assess personality are often inconsistent with the research that linked personality to job performance in the first place. Although his article did go on to say that personality measures often offer good predictive validity,

Hogan (2005) noted additional concerns. He wrote that: there is little agreement in the field of personality research regarding the agenda of the discipline; personality research often evidences a lack of concern for validity; and research in personality has often been poorly done (Hogan, 2005).

Hough and Oswald (2005) offered a response that agreed with many of the criticisms of Murphy (2005) and Hogan (2005), while assuming a different perspective. They contended that: different agendas in personality research actually allow for greater productivity; most recent personality research had been conducted in areas where validity was of the utmost concern; low validities have been shown to increase as predictors are more thoughtfully matched with criteria; and that continuing such research will result in better theories regarding the links between personality constructs and job performance (Hough & Oswald, 2005). While they advanced a respectable rebuttal to many criticisms, the matter was far from settled.

In fact, Murphy was co-author of a subsequent publication that ignited a considerable debate regarding this topic (Morgeson et al., 2007a; Morgeson, et al., 2007b; Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett, & Christiansen, 2007). With their report of a panel discussion held at the 2004 Society for Industrial and Organizational Psychology (SIOP) conference, Morgeson et al. (2007b) noted several potential problems with the use of personality inventories in selection contexts. They reached several conclusions, including that: personality tests often have low validity for predicting overall job performance; many published personality tests should not be used; customized measures of personality that are clearly job-related (face valid) may be more

appropriate; faking on self-report personality measures should be expected and may be desirable as a form of social adaptability; corrections for faking do not improve validity; and alternatives to self-report measures of personality should be sought (Morgeson et al., 2007b).

Extensively citing previous literature, Ones et al.'s (2007) response to Morgeson et al. (2007b) noted in part that: Conscientiousness and its facets have evidenced similar validity to other frequently used methods such as assessment centers, structured interviews, biodata and situational judgment tests; various groups of personality predictors offer predictive ability for multiple occupations; it is unclear why they expect less validated measures of personality to offer higher predictive ability; although faking is always a concern with such tests, evidence supporting the criterion-related validity of personality tests suggests that the impact of faking is overrated; proposed alternatives to the traditional approach of correcting for faking have met with some success; and that selection decisions made without including personality characteristics would be inherently deficient. Tett and Christiansen (2007) also offered a refutation, stating in part that: the validity of personality tests is often underestimated due to methodological issues, yet reaches useful levels when care is taken in applying them under appropriate conditions that are based on well-constructed theory; personality measures often offer useful incremental validity when combined with other predictors; although faking may attenuate personality test validity, enough variance remains for useful prediction; and that faking as a socially desirable behavior has not been empirically demonstrated enough to be tolerable in selection contexts.

A second publication from Morgeson et al. (2007a) offered several rejoinders, including statements that: they had referred only to job performance criteria (deemed by them to be the most important criteria, as opposed to contextual aspects of work) with their previous article, where validities remain disappointingly low; and they reinforced their belief that faking remains an issue of great importance, as the responses from Ones et al. (2007) and Tett and Christiansen (2007) both seemed to acknowledge (although to varying degrees). However, their second article also eventually conceded that personality is likely to be correlated with variables of interest to organizations and researchers and that the authors do not believe that such measures inventories lack relevance to the understanding of behavior in the workplace (Morgeson et al., 2007a).

Indeed, when one considers the multidimensional nature of work today, which incorporates not only task performance but also contextual aspects such as organizational citizenship behaviors (OCB) and counterproductive work behaviors (CWB), the role of personality in selection becomes even more apparent (Gonzalez-Mulé, Mount, & Oh, in press). For instance, a recent meta-analysis from Gonzalez-Mulé et al. (in press) reports that FFM traits are more important than GMA in predicting contextual aspects of work such as CWB, equally important to GMA in the prediction of contextual aspects of work such as OCB, and less important than GMA in the prediction of task aspects of work and overall performance.

It seems that organizations are following suit, as Rothstein and Goffin (2006) noted in their comprehensive review of the topic that organizations are increasingly using personality measures for selection purposes. For example, Heller (2003) reported that

over 30% of American companies use personality tests to screen job applicants. Additionally, Erickson (2004) reported a survey conducted by the Society for Human Resource Management (SHRM) stating that over 40% of Fortune 100 companies use personality testing for applicants at varying levels of employment. When one observes such current trends in I/O Psychology, it seems that (although the practice may never be without its detractors) the use of personality tests in selection contexts is likely to continue.

Faking and Personality Assessment

The debate over the use of personality measures in selection contexts may be fairly easily dismissed by accepting that this practice possesses at least some utility to organizations (whether referring to its predictive ability of task performance, or simply that of contextual behaviors), as acknowledged to varying degrees in the previously outlined research (Avis et al., 2002; Dudley et al., 2006; Gonzalez-Mulé et al., in press; Judge et al., 2013; Li et al., 2014). Considering this (along with the increasing use of such measures), the discussion of the potential impact of faking remains an important one. Before beginning the discussion regarding faking and its impact on personality measures in selection contexts, it is important to begin by defining the construct and reviewing the many terms used in the literature to refer to it.

In the past, some have included unconscious behaviors that distort the truth to promote a positive impression, although most current researchers use definitions that imply or explicitly require intent (MacCann, Ziegler, & Roberts, 2011; Paulhus, 1984). Vispoel and Tao (2013) define *socially desirable responding* as a tendency (unconscious

or willful) to respond to items to make appear good rather than be truthful. Hogan, Barrett, and Hogan (2007) defined impression management as a process in which one controls their behavior during social interaction, which includes responding to inventories. Ones, Viswesvaran, and Reiss (1996) suggested that the term *response distortion* serves as an umbrella term under which fall those previously listed and most others, such as: *self-enhancement*, *claiming unlikely virtues*, *denying common faults*, *self-presentation*, *faking*, etc. Response distortion has been defined as distorted responses intended to create favorable self-presentation and outcomes (Rosse, Stecher, Miller, & Levin, 1998).

The list of terms used in the literature to describe such behavior is lengthy indeed, however, the most commonly used appellation seems simply to be *faking* (Griffith & Peterson, 2006; Griffith & Peterson, 2011; Hogan et al., 2007; Kuncel & Borneman, 2007; Morgeson et al., 2007; Tett & Christiansen, 2007; Ziegler et al., 2011). Kuncel & Borneman (2007) define faking as a conscious attempt to present misleading and deceptive information about one's personality, interests, experiences, past behaviors, and attitudes in order to influence others. Simplifying, Holden and Book (2011) offered a definition for faking as a self-report containing intentional misrepresentation. They elaborated that the three key features of faking are the implications of: intent, deception, and orientation toward others (Holden & Book, 2011). This definition is comprehensive enough to subsume most of those previously mentioned, as well as to include most relevant research on the topic. Notwithstanding, many of these synonymous labels will be used throughout this paper in an effort to discuss prior research in its original terms.

To be thorough, I believe that a discussion of the faking literature must include two important aspects: the nature of designs used to study faking, and the divergent viewpoints that have resulted from such research. The literature generally cites two methods of faking research: directed faking studies conducted in lab settings, and applicant studies conducted in real-world selection contexts (Viswesvaran & Ones, 1999). Upon review of the reported results from such methods, it seems that most researchers generally agree that faking on personality measures occurs to some degree (MacCann et al., 2011; Morgeson et al., 2007). For various reasons, however, some believe this to be a legitimate concern when making selection decisions, while others argue that it does not represent a significant problem (Hogan, 2005; Hogan et al., 2007; Morgeson et al., 2007; Ones et al., 1996). I will begin my discussion of faking with a review of the research designs typically employed.

Directed-Faking in Laboratory Studies

Directed faking research is an approach wherein participants (in within- or between-subjects designs, often conducted in laboratory settings) complete the same personality measure: in one condition participants are instructed to be honest, and in another they are instructed to fake. An evaluation of score differences between conditions provides an indicator of the maximum degree of fakability for that measure (Viswesvaran & Ones, 1999). In their meta-analysis regarding such research, Viswesvaran and Ones (1999) found that all five factors of the FFM were equally susceptible to faking in between-subjects designs, yet evidenced increased variability and larger effect sizes in within-subjects designs.

This suggests that faking research may best be conducted using within-subjects designs, which eliminate reductions in variability due to analyzing group-level differences rather than focusing on potentially impactful changes at the individual-level (Mesmer-Magnus & Viswesvaran, 2006; Viswesvaran & Ones, 1999). In addition, within-subjects designs offer greater statistical power (Mesmer-Magnus & Viswesvaran, 2006). Nevertheless, limitations to the within-subjects approach include the threat of testing effects as a result of participants completing the same measure twice, varying from several months to a week or even just a day later (Costa & McCrae, 1992; Kuncel, & Borneman, 2007; Mesmer-Magnus & Viswesvaran, 2006). Of note, individuals in within-subjects designs substantially altered their scores the most on Conscientiousness and Emotional Stability scales (the two scales most often reported as valid predictors of performance) when instructed to do so (Viswesvaran & Ones, 1999).

Critics of the directed-faking approach argue that it might be methodologically unsound (Abrahams, Neumann, & Githens, 1971; Hogan et al., 2007; Smith & Robie, 2004). Consider Snell, Sydell, and Lueke's (1999) theory regarding faking, echoed by Goffin and Boyd (2009), which proposes that faking behavior has two components: the motivation to fake and the ability to do so. These models parallel Hogan's (1991) theory of impression management as being comprised of the desire to distort responses and the ability to do so. Smith and Robie (2004) cite Hogan's (1991) theory in their publication while suggesting that such laboratory research only evaluates one aspect of faking because it holds the desire to fake constant by instructing everyone in a particular condition to do so. Because they inherently assume that all applicants will choose to engage in faking behavior when motivated to do so, such research designs analyze only

individual differences in the ability to fake. However, it has been extensively reported that the desire and ability to fake vary per individual (Barrick & Mount, 1996; Hogan et al., 2007; McFarland & Ryan, 2000; Morgeson et al., 2007; Smith & Robie, 2004; Viswesvaran & Ones, 1999).

As Smith and Ellingson (2002) state, such manipulations exaggerate the effects of such behavior beyond what would be expected in real-world contexts. Reports that laboratory and field studies differ in effect sizes and variability (with lab studies evidencing greater levels of both) when making comparisons between faking (applicant) and non-faking (incumbent) groups further substantiate such skepticism (Hough, 1998; Viswesvaran & Ones, 1999). Certainly, it stands to reason that greater levels of faking are obtained in studies comparing participants in a condition that instructs everyone to fake with another condition in which everyone is instructed to be honest, as opposed to studies comparing one condition in which applicants can choose whether to fake or not at their own discretion (most may not) with a condition in which all are presumably responding honestly. Therefore, critics of this approach assert that such research is unlikely to result in conclusions that will generalize to real-world selection contexts (Abrahams et al., 1971; Hogan et al., 2007; Smith & Ellingson, 2002; Smith & Robie, 2004).

Applicant Research in High-Stakes Contexts

In applicant research, applicants' scores on personality tests are often referenced against validity scales and/or similar scores from non-applicant samples (usually job incumbents). Higher scores from the applicant sample are thought to indicate faking, with the difference representing a measure of such behavior at a real-world level

(Viswesvaran & Ones, 1999). As Smith and Robie (2004) report, consistent differences in mean scores between applicants and incumbents have been found in such studies. In one such study using a between-subjects design, Rosse et al. (1998) found: significantly greater response distortion for applicants when compared to incumbents, substantial variance among applicants regarding the levels of distortion they exhibited, and negative impact on hiring decisions due to such response distortion.

In another study using a within-subjects design, Barrick and Mount (1996) found varying levels of response distortion amongst applicants, but also reported that such distortion did not affect the predictive ability of the factors of Conscientiousness and Emotional Stability. While both studies agree that applicants differ in their level of response distortion, the dissimilar conclusions regarding predictive ability may have been due to methodological differences. Aside from using within- and between-subjects designs respectively, Rosse et al. (1998) analyzed the impact of faking at the individual-level by examining changes in rank-order that occurred due to faking, while Barrick and Mount (1996) analyzed mean performance outcomes. Additionally, (like many studies regarding faking) both studies mentioned here used scales of validity to estimate levels of applicant faking, which many have suggested (elaborated upon later in this paper) are unreliable measures of such behavior (Barrick & Mount, 1996; Ispas et al., 2014; Ones et al., 1996; Rosse et al., 1998)

Continuing, while a method that utilizes real-world applicants undoubtedly provides an inherent boost in internal validity and generalizability, the results do not present an airtight case. For instance, following the notion that individuals who had been

previously rejected for a job would be motivated to improve their scores on a second application, Hogan et al. (2007) found no meaningful changes from time one to time two in one within-subjects study using real-world applicants. In fact, the authors found that applicants' scores decreased from time one to time two as often as they improved (Hogan et al., 2007). However, in a second study Hogan et al. (2007) also found evidence that individual differences in faking were moderated by scores for social skills and social desirability, further substantiating that faking behavior varies per individual.

Additionally, Hogan et al. (2007) presented a third study that addressed the concern that conclusions could not be reasonably drawn from the previous two as to whether or not faking had equally occurred during the first application as well as the second. Using a between-subjects approach, the authors found that the scores of applicants completing a personality inventory for research purposes only (presumably with no motivation to fake) and those of applicants completing the measure for selection purposes (even for a second time, after having been previously rejected) did not significantly differ (Hogan et al., 2007). However, it should be noted that the internal validity of this third study suffers to a degree from its reliance on a between-subjects design, as opposed to the within-subjects design of the first two. Additionally, all three studies relied mostly on analyses of performance at the group-level, which fail to maximize insight into potential changes due to faking that occur at the individual level.

As the research outlined above illustrates, the case for differential faking at the individual level seems to be supported with the improvement in validity using the applicant research approach. This would be expected to increment even further when

combined with within-subjects designs. Nonetheless, mixed results such as those described above warrant the continued investigation of applicant research. Considering previous recommendations and literature, it seems that combining the within-subjects design suggested by Ones and Viswesvaran (1999) and a repeated-measures, between-subjects design that controls for testing effects and uses real-world applicants may represent a worthy form of investigation. As will be outlined in subsequent sections, the current research attempts to follow this approach.

Faking: Insignificant Problem or Legitimate Concern?

The previous sections lead directly into a discussion of the divergent perspectives held by researchers regarding the impact of faking on personality measures in selection contexts. The reasons behind the perspective that faking represents an insignificant problem include: Morgeson et al.'s (2007b) suggestion that such behaviors actually represent desirable manifestations of social adaptability, which follows logic similar to Hogan's (2005) assertion that it is difficult to distinguish faking from socialized behavior; Hogan et al.'s (2007) point that the base rate of applicants who engage in faking is small; and Ones et al.'s (1996) conclusion that social desirability does not affect criterion-related validities. In the following sections, I will attempt to provide points in support of the perspective questioning the significance of the problem, as well as counterpoints that demonstrate why faking remains a legitimate concern.

Is Faking Socially Adaptive?

Morgeson et al.'s (2007b) assertion, which echoed Hogan's (2005), stems from the idea that everyday life is comprised of a compromise between authentic behavior and

faking. This view asserts that social interaction is made up of people behaving in a certain manner to impress upon others a positive image of themselves. It maintains that such behaviors are basically true, only more or less emphasized depending on the current circumstances and goals of the individual (Schlenker & Weigold, 1992). Murphy, in Morgeson et al.'s (2007b) article, went so far as to describe such behavior (wherein one says what they think they should rather than what they would really like to say) as analogous to civilization. Murphy went on to make the point that it should be more perplexing when people act in ways that are not socially adaptive in situations in which most people act as expected (Morgeson et al., 2007b). Smith and McDaniel (2011) extend the argument, writing that it seems reasonable to expect successful fakers (those who are hired) to perform on the job consistently with their faking in the selection process because of the knowledge of job demands and requirements required to fake successfully.

Tett and Christiansen's (2007) criticism, that this perspective lacks empirical support, has been previously noted. Furthermore, Johnson and Hogan (2006) wrote that it is in the best interest of both employees and organizations to hire people with characteristics right for a particular job. Additionally, as evidence supporting individual differences in variability and degree of faking behavior suggests, some individuals may fake in a manner consistent with their everyday levels of adaptability, while others may exaggerate to extreme degrees that they are incapable of maintaining in real-world contexts (Barrick & Mount, 1996; Hogan et al.'s, 2007; McFarland & Ryan, 2000; Morgeson et al, 2007b; Rosse et al., 1998; Smith & Robie, 2004; Viswesvaran & Ones,

1999). Therefore, the aforementioned reasonable expectation advanced by Smith and McDaniel (2011), of job performance matching faking in selection contexts, may not be reasonable for all fakers.

Certainly, faking may be viewed as initially adaptive if it ends in obtaining a desired position, however, doing so may also decrease that individual's job satisfaction and chances for success if it results in poor fit (Holland, 1997). Therefore, even if one accepts the view of a certain level of faking as being socially adaptive, one must also concede that extreme levels of such behavior may become problematic. Consider the results from Rosse, Levin, and Nowicki's (1999) study, which found that faking may be positively related to negative social behaviors such as misleading customers and making impossible promises to customers, as well as negatively related to positive behaviors such as listening carefully to customers and suggesting products that fit customers' needs (Rosse et al., 1999). While such behavior may positively impact an organization's bottom line and employee sales commissions, it seems rather difficult to construe it as socially adaptive.

What is the Prevalence of Faking?

Hogan et al. (2007) cited multiple sources in making the point regarding what was referred to as a minimal base-rate of faking on job applications. Ellingson, Sackett, and Connelly's (2007) study supported such an assessment, reporting score increases due to response distortion of just less than one-tenth of one standard deviation. However, a closer review of the literature cited in Hogan et al.'s (2007) article reveals that several of these publications simply reported finding smaller differences (as opposed to negligible

differences) in applicant research when compared with directed-faking studies (Hough, 1998; Hough & Ones, 2001). In addition, such results are to be expected when one considers the previously mentioned limitations of the directed-faking approach, which serve to exaggerate differences between conditions at the group-level (Smith & Ellingson, 2002).

Furthermore, contradictory results have been reported with relative frequency. For example, after assessing faking behaviors of recent job applicants using a survey Donovan, Dwight, and Hurtz (2003) reported a base-rate for faking at approximately one-third (over 30% for many of the assessed behaviors) of applicants, and suggested that this may actually be an underestimation. In addition, the authors went on to report that 50% of respondents admitted exaggerating qualities (such as dependability and reliability) associated with Conscientiousness (Donovan et al., 2003). Relatedly, in a within-subjects study, Griffith et al. (2007) reported that 31% of applicants elevated their Conscientiousness scores when the estimation was based on a 95% confidence interval for an individual's applicant score as it related to their honest score (Griffith et al., 2007). This study also found that 22% of applicants would still have been categorized as fakers when a more conservative confidence interval was used (Griffith et al., 2007).

Rosse et al.'s (1998) between-subjects study found that applicants scored on average 0.69 standard deviations higher than incumbents on FFM facet-level scales, and concluded that these levels were similar to those in directed-faking studies. Citing multiple sources involving various study designs, MacCann et al. (2011) wrote that the general consensus is that about one in four individuals fake in high-stakes situations.

Although such vernacular is somewhat open to idiosyncratic interpretation, widespread reports of such findings make Hogan's (2005) assertion (that the base-rate of applicant faking is minimal) seem questionable.

Providing further reason to remain concerned over potential faking, as previously mentioned, individual differences in degree and style of faking have been evidenced (Barrick & Mount, 1996; Hogan et al.'s, 2007; McFarland & Ryan, 2000; Morgeson et al., 2007b; Rosse et al., 1998; Smith & Robie, 2004; Viswesvaran & Ones, 1999). Therefore, even if base-rates were minimal (which seems arguable), the finding that individual differences in faking behavior exist suggests that certain applicants have the potential to fake to impactful degrees, making faking in selection settings a concern for organizations and researchers alike.

For instance, in a within-subjects design (counterbalanced between-subjects for order effects) McFarland and Ryan (2000) found that individuals low in Conscientiousness and Emotional Stability faked to a greater extent when directed to do so. Similarly, Birkeland, Manson, Kisamore, Brannick, and Smith's (2006) meta-analysis of between-subjects faking studies in real-world contexts found that applicants faked most for Conscientiousness and Neuroticism, with significant (but much lower) effects for Extraversion and Openness as well. Although Ones et al.'s (1996) earlier meta-analysis reported positive correlations between social desirability scales and both Conscientiousness and Emotional Stability, the mixed results still represent cause for concern.

As these two constructs are the most cited factors of the FFM as predictive of work outcomes, any evidence of the potential for such faking should be viewed as highly relevant in selection contexts (Barrick & Mount, 1991; Hogan & Holland, 2003; Hurtz & Donovan, 2000; Salgado, 1997). Indeed, in summarizing several of the aforementioned studies (along with many others), Tett, Anderson, Ho, Yang, Huang, and Hanvongse (2006) concluded that the degree of applicant faking is high enough to warrant concern regarding selection decisions, and that applicants tend to fake on measures of traits most believe to be related to performance.

Does Faking Affect the Predictive Validity of Personality Measures?

Finally, the research regarding the effects of faking on criterion-related validities also offers mixed results, and may even be somewhat misguided. In a between-subjects laboratory study, Mueller-Hanson, Heggstad, and Thornton III (2003) found that the relation between predictor and criterion in a control (non-faking) condition was relatively constant (.26 to .20) from the bottom third of the distribution to the top third in their within-subjects study. However, in an incentivized condition (that simulated real-world motivation to fake) this relation changed dramatically, nearly disappearing at the top third while moving from .45 to .07 (Mueller-Hanson et al., 2003). These differences between groups (although not statistically significant) suggested to the authors an impact on criterion-related validity due to faking (Mueller-Hanson et al., 2003). Additionally, Converse, Peterson, and Griffith (2009) later reported differences between faking and honest validities for a measure of Conscientiousness, and for multiple predictor models that included Conscientiousness, resulting from their faking simulation.

Alternatively, Hough, Eaton, Dunnette, Kamp, and McCloy (1990) reported that validities remained stable regardless of possible distortion in their directed faking study using a between-subjects design. However, nearly two-thirds of the validity comparisons (between accurate and overly desirable responders) in this study evidenced greater or equal validities between groups, although slightly less than one-third proved to be statistically significant (Hough et al., 1990). Further, in their meta-analysis of social desirability studies, Ones et al. (1996) found that removing the effects of socially desirable responding did not affect criterion-related validities and concluded that faking is not as great a problem as researchers had previously believed. However, their assertion that scales of social desirability might actually represent substantive trait variance suggests that such scales are ineffective as measures of faking behavior and seem to contradict their conclusion regarding faking as an insignificant problem (MacCann et al., 2011; Ones et al., 1996). Similarly, Schmitt and Oswald (2006) later reported results from their simulation that correcting for faking did not impact the predictive ability of non-cognitive measures on mean performance. However, they also noted that faking might cause problems at the individual-level in terms of selection rates with the possible displacement of honest respondents (Schmitt & Oswald, 2006).

In support of this notion, it has been reported that changes in rank-order (due to faking) at the upper end of the distribution can occur without impacting measures of criterion-related validity, thus resulting in an impact due to faking that is not manifest when one considers only a criterion-related validity coefficient (Christiansen, Goffin, Johnston, & Rothstein, 1994; Rosse et al., 1998). Therefore, lack of changes evidenced in criterion-related validities need not imply that faking has no impact on selection

decisions and subsequent performance at the individual-level. In fact, Griffith and Peterson (2011) eventually concluded that an overemphasis on empirical questions (such as the one regarding faking as a threat to criterion-related validities) has resulted in a lack of solid theory development regarding applicant faking, leaving advances in efforts toward faking control and detection difficult. More on the potential impact of faking at the individual-level follows in the subsequent section, which discusses faking as it relates to selection rates.

The Impact of Faking on Selection Rates and Hiring Decisions

As one can clearly see, faking behavior is an area that commands great interest (and results in divergent perspectives) from researchers in the field of I/O Psychology. Recent research regarding faking on personality measures has evidenced the potentially harmful impact the behavior may have on honest responders at various selection rates (Christiansen et al., 1994; Peterson, Griffith, & Converse, 2009; Mueller-Hanson et al., 2003; Rosse et al., 1998; Winkelspecht, Lewis, & Thomas, 2006).

Selection systems typically come in two basic forms: those that use select-in strategies and those that use select-out strategies. Nearly all of the faking research has focused on the former, with which this discussion will begin. First, it is worth repeating that the faking literature often reports and cited that criterion-related validities have often not been attenuated due to faking behaviors (Barrick & Mount, 1996; Hough et al., 1990; Ones et al., 1996). While some recent research has contradicted these reports, of greater concern may be the finding that faking has often been found to cause changes in rank-order that profoundly affect individuals, yet do not impact criterion-related validity

coefficients (Christiansen et al., 1994; Converse et al., 2009; Komar, Brown, Komar, & Robie, 2008; Mueller-Hanson et al., 2003; Rosse et al., 1998). Findings such as these (discussed in the following section) provide a direct impetus for research into faking as it pertains to select-in decisions.

Faking and Select-In Hiring Decisions

Select-in hiring decisions involve top-down systems in which applicants are rank-ordered based on test scores, and those with the highest scores are selected until all available positions have been filled (Mueller-Hanson et al., 2003). Christiansen et al. (1994) published the results from one between-subjects study regarding faking as it affects select-in procedures, which evaluated correcting 16PF (a well-known inventory that was developed to assess normal-range personality) scores for faking at various selection rates (Cattell & Mead, 2008). The authors found that more discrepant hires (candidates hired on the basis of uncorrected 16PF scores, that would not have been hired with corrected scores) for upper-level supervisory positions were found at lower selection rates.

However, they also concluded that the lack of change in criterion-related validities supported the notion that faking may not be a serious threat to the validity of personality measures (Christiansen et al., 1994). Ellingson, Sackett, and Hough (1999) examined the effects of score corrections for faking on different selection rates using Conscientiousness as a predictor. While their within-subjects design (allowing them to analyze score changes at the individual-level) evidenced that corrected scores were closer to honest scores, ultimately they obtained mixed results. The corrections resulted in

higher selection proportions of non-fakers in only some situations, while in others the opposite result occurred (Ellingson et al., 1999). However, similar to Christiansen et al. (1994), the authors also found that larger selection rates were associated with lower proportions of hired fakers (Ellingson et al., 1999).

The results of Rosse et al.'s (1998) study led the authors to suggest that research into faking should shift focus to hiring decisions rather than predictive validity. The authors rank-ordered job applicants who responded to the NEO-PI-R (a revision of the previously mentioned NEO-PI, which is discussed in more detail in the measures section of this paper) according to their scores on the Conscientiousness scale (Costa & McCrae, 1992; Rosse et al., 1998). They found that the average level of response distortion in hired applicants would have been at least one standard deviation above the mean at selection rates of 25% or less (Rosse et al., 1998). Similar to the previously mentioned studies, with increasingly smaller selection rates the effects were even more dramatic. The authors reported that more than half of the hired applicants would have had extreme response distortion scores if hiring were limited to the top 10%, whereas seven out of eight of those hired would have had similar scores if the selection rate were limited to the top 5% of applicants (Rosse et al., 1998). Although such results provide some cause for concern, questions remain. The mixed results, use of between-subjects design that precludes analyses of faking at the individual level, and/or measuring faking using validity scales that have come to be regarded as ineffective measures of such behavior require further examination of such effects (MacCann et al., 2011).

Converse et al.'s (2009) simulation extended this research by examining the effects of single-predictor (Conscientiousness alone) and multiple-predictor (Conscientiousness with multiple predictors) models. They found displacement of honest responders using both models, and concluded that multiple-predictor models were less susceptible to the negative effects of faking (Converse et al., 2009). The same three authors also reported on a similar study that involved actual participants in a within-subjects design, comparing Conscientiousness alone and Conscientiousness with cognitive ability, at multiple selection rates (Peterson, Griffith, Converse, 2009). Echoing the findings of Christiansen et al. (1994) and Rosse et al. (1998), the effects of faking (as identified by a change score from honest to applicant conditions for each participant) when using Conscientiousness as a lone predictor were more pronounced at lower selection rates (Peterson et al., 2009).

At a selection rate of 10%, over 40% of the hired applicants would have been fakers, while at a rate of 30% the percentage of hired fakers only dropped to slightly less than 30 (Peterson et al., 2009). Additionally, they found that nearly all (100%, 100%, and 97%, respectively) of the fakers at the three selection rates would not have been hired when using their honest scores (Peterson et al., 2009). Further, adding cognitive ability as a predictor (using multiple combination techniques) did not significantly alter the percentage of fakers hired at any of the selection rates, although it did reduce the percentage of discrepant hires at all three selection rates (Peterson et al., 2009). Ultimately, the authors concluded that the use of multiple predictors can result in reducing the negative impact of faking, but the lack of statistically significant reductions

in most situations suggested that concern over faking could not be disregarded (Peterson et al., 2009).

Similar findings have been replicated often throughout the literature. As previously mentioned, Mueller-Hanson et al. (2003) found decreased levels of criterion-related validity at the high end of the distribution for the faking group in their between-subjects study in a lab setting with an incentivized condition incorporating realistic motivation to fake. The authors also found that those in the faking group evidenced lower mean performance, yet were more likely to be selected as selection rates grew smaller (Mueller-Hanson et al., 2003). While acknowledging that correcting for faking made little difference on overall group-level analyses of criterion-related validity, Schmitt and Oswald (2006) also reported that the selection rate accounted for 23% of the variance in mean performance, such that smaller selection rates led to greater differences in performance between faking-corrected versus non-corrected groups in their simulation.

Using the NEO-PI-R, Winkelspecht et al. (2006) found that individuals in a directed-faking condition were over-represented at the top of score distributions in their between subjects study, and that these percentages became increasingly disproportionate with smaller selection rates. Using a measure of Conscientiousness that correlates with that scale from the NEO-PI-R, Griffith et al. (2007) found that faking behavior in applicants resulted in rank-order changes at the top of the distribution across three selection rates in a within-subjects study using real-world applicants, with the most significant changes being at the smallest selection rate examined.

The extensive amount of literature on the topic makes the argument that faking affects individual-level select-in decisions and that smaller selection rates result in greater deleterious effects of such behavior. However, these studies often suffered from limitations as a result of relying on between-groups designs, directed faking, the use of validity scales, or the use of simulations rather than human participants. To evaluate most efficiently the real-world effects of faking behavior, it may be necessary to adopt a within-subjects design involving actual job applicants whose faking behavior is measured with a reliable form of analysis. This will allow for the identification of individual-level response differences (indicating faking) between conditions free of possible differences due to group characteristics. In addition, such a design will represent true levels of faking prevalence, unlike directed-faking studies (that exaggerate such effects) or studies that have relied on inefficient measures of faking behavior (such as validity scales). This study will add to the literature by attempting to meet these demands.

Faking and Select-Out Hiring Decisions

The extant literature often suggests that the evidence of changes in rank-order at the high end of distributions suggests that select-out strategies may remain less affected by faking behavior than their select-in counterparts (Christiansen et al., 1994; Mueller-Hanson et al., 2003; Rosse et al., 1998). In select-out hiring strategies, applicants are also rank-ordered on the basis of test scores, but a minimum qualification threshold (also known as a cut-score, below which applicants are not considered for employment) is established (Berry & Sackett, 2009; Mueller-Hanson et al., 2003). This reduces the size of the applicant pool by eliminating those found to be the worst performers (Mueller-

Hanson et al., 2003). In addition, this method is thought to address the concern that top-down (select-in) selection systems often cause honest respondents to be displaced by fakers (Berry & Sackett, 2009). Mueller-Hanson et al. (2003) briefly discussed the potential effects of faking on select-out strategies, concluding that such an approach will result in effectively removing low performers as the results of their study indicated that criterion-related validities were maintained at the lower end of the faking distribution. They suggested that personality tests are perhaps best used in select-out, rather than select-in, contexts (Mueller-Hanson et al., 2003).

Berry and Sackett (2009) offer what is (to my knowledge) the only dedicated analysis of the effects of faking behaviors in select-out contexts heretofore reported upon in the extant literature. Their study compared two methods of applying cut-scores as they were affected by faking. According to Berry and Sackett (2009), the applicant-data-derived (ADD) method uses the test scores of applicants (which may include individuals motivated to fake responses) to establish the cut-score. The non-applicant-data-derived (NADD) method uses the test scores of non-applicants (presumably with no motivation to fake) to establish the cut-score (Berry & Sackett, 2009). The critical difference between the two methods is that different candidates may comprise the group above the cut-score with the ADD strategy, depending on whether faking behaviors occur (Berry & Sackett, 2009). Berry and Sackett (2009) suggest that this situation results in what they deemed a tradeoff.

Across various selection rates, the authors found that mean performance was always lower with the NADD approach. The authors concluded that adopting the NADD

strategy often results in higher passing rates than desired by the organization (Berry & Sackett, 2009). In addition (as one might expect after reviewing the literature regarding faking and select-in strategies), the differences in average performance between the two cut-score strategies became larger as selection rates decreased (Berry & Sackett, 2009). The suggested tradeoff is realized when considering that the authors also found that adopting the ADD approach instead resulted in more displacement of deserving applicants (Berry & Sackett, 2009). Again, the size of the selection rate evidenced similar effects, with greater levels of displacement at smaller selection rates (Berry & Sackett, 2009). In summary, while mean performance may increase due to the ADD strategy, the NADD strategy may allow an organization to avoid more unfair displacement. Ultimately, the authors suggested that organizations consider what they value most from the selection process, and choose their select-out method accordingly (Berry & Sackett, 2009).

Previous Approaches Used to Address Concerns Regarding Potential Faking

As the select-out strategy effectively represents an attempt to control for some of the deleterious effects of faking that occur during select-in contexts, the previous discussion functions well as a segue to the following section which discusses various approaches that attempt to address the problem of faking. As described above, vast amounts of research and literature in the field of I/O Psychology have been devoted to concerns regarding potential faking on measures of personality. Certainly, the potential for an individual to misrepresent oneself in order to gain a desired outcome affects the very core of I/O Psychology, which is in large part based on the accuracy of

psychometric predictions. In addition, as the discussion on faking and selection rates illustrates, fakers may often be quite successful in obtaining such desired outcomes (Berry & Sackett, 2009; Christiansen et al., 1994; Converse et al., 2009; Mueller-Hanson et al., 2003; Rosse et al., 1998; Winkelspecht et al., 2006).

Therefore, as one might have guessed, a multitude of approaches have been employed in an effort to attend to this problem. According to Kuncel and Borneman (2007), these efforts fall mainly under two categories. *Control efforts* attempt to suppress or eliminate faking altogether, while *detection efforts* attempt to identify those individuals likely to have engaged in such behavior (Kuncel & Borneman, 2007). While a discussion of these methods is important to the understanding of contemporary thinking regarding faking, it is important to note that they have mostly met with minimal success. Such results are perhaps the consequence of notable limitations respectively diminishing the efficacy of each of these methods. I will begin with a discussion of control efforts.

Methods that Attempt to Control or Eliminate the Problem

Efforts to control or eliminate faking have typically involved instructions or warnings against faking, or item formats designed to make faking more difficult (Kuncel & Borneman, 2007). I will start my discussion of faking control efforts with a review of the literature regarding warnings against faking. The idea behind the utility of using warnings against faking is often based on the notion that lower scores on non-cognitive measures by those given warnings are a result of a reduction in faking (Dwight & Donovan, 2003). While some success in reducing faking has been found using warnings, it has been reported that this type of control effort may be contingent upon certain factors

(such as item transparency and warning type), and even then often results in only minimal effect sizes (Dwight & Donovan, 2003). According to Dwight and Donovan's (2003) review of the warning literature, two types of warning against faking are typical: those suggesting that individuals engaging in faking behavior will be *identified*, and those suggesting that there will be *consequences* for faking behavior.

One study examining the consequences approach found that the majority of applicants that had initially submitted invalid personality profiles (MMPI-2 profiles with elevated L or K scale scores, which are two types of validity scales that will be elaborated upon in the following section) later submitted valid profiles after having been presented with instructions regarding the possible invalidation of their results due to dishonesty (Butcher, Morfitt, Rouse, & Holden, 1997; Butcher & Tellegen, 1966). In fact, according to Dwight and Donovan's (2003) review of warnings research, many studies have reported significant effects for various warnings approaches. However, the authors also noted a sporadic pattern of results and called into question their practical significance (Dwight & Donovan, 2003). First, the authors noted that warnings involving consequences had larger effects than those offering only the threat of identification (Dwight & Donovan, 2003). In addition, after weighting the effect sizes according to sample size, they found that all warnings had an overall mean *d* of just .23 (with nearly half of the analyzed studies showing almost no effect, and less than 20% evidencing a *d* greater than .30 in the desired direction), suggesting a weak effect of warnings' ability to reduce faking (Cohen, 1988; Dwight & Donovan, 2003).

Further, although these modest effect sizes may indeed represent improved validity compared with tests given without warnings, these effects were mostly found at the group-level (Dwight & Donovan, 2003). Thus, while such efforts serve to reduce the overall success of faking at the individual-level, several fakers may still be able to beat the system and be hired when involved in top-down selection processes (Dwight & Donovan, 2003). This variance in outcome at the individual-level may be the result of warnings being dependent on participants' naïve beliefs regarding yet to be developed measures for faking detection and/or the development of such measures, which are likely to vary per individual (Kuncel & Borneman, 2007).

A subsequent article by Fan, Gao, Carroll, Lopez, Tian, and Meng (2012) examined warnings given early on in the testing process to those identified as potential fakers and offered some promise to this approach, but was also accompanied by several practical issues. The authors found that participants receiving a warning message lowered their subsequent personality scores in comparison to those in a control condition (Fan et al., 2012). However, while faking was reduced as a result of this method, the authors conceded that it was not eliminated completely (Fan et al., 2012). In addition, the authors reported that alternative explanations for the evidenced score reductions (such as regression toward the mean) could not be ruled out (Fan et al., 2012).

Extending the elaboration on methods that attempt to control or eliminate faking, the item format method typically suggested involves a paired-comparison, forced-choice approach. In this style of testing, participants must choose between two or more options of similar desirability but differential validity (Christiansen, Burns, & Montgomery,

2005). This often involves balancing statements within an item for social desirability such that it becomes difficult to respond on this basis alone, thereby reducing the potential for intentional distortion (Converse, Oswald, Imus, Hedricks, Roy, & Butera, 2008). An example of such an item asks the test taker to choose what is most like them between the following options: “Once I give priority to a project, I follow it through,” or “I’m usually the first person to strike up a conversation with strangers” (Converse, Oswald, Imus, Hedricks, Roy, & Butera, 2006, p. 268). The first option is a desirable item reflecting Conscientiousness, while the second is a desirable item reflecting extraversion (Converse et al., 2006).

In one study, Jackson, Wroblewski, and Ashton (2000) found that participants could increase their scores on an integrity test by nearly one standard deviation when instructed to fake-good on a normative (single-statement items, Likert-scored) test. Participants in the same study were only able to increase their scores by less than one-third of a standard deviation when responding to a paired-comparison, forced-choice measure (Jackson et al., 2000). Similarly, Martin, Bowen, and Hunt (2002) found that the test form (forced-choice vs. normative) moderated the relationship between degree of faking and test instructions (honest vs. faking), such that significantly more faking was evidenced for the faking group over the honest group on the normative test as compared to the difference between those groups on the forced-choice measure (for which no significant difference was evidenced).

The optimism inspired by such findings has, nevertheless, been somewhat tempered by further research. For example, Heggstad, Morrison, Reeve, and McCloy

(2006) found that participants were able to raise their scores on Conscientiousness scales on both normative and forced-choice measures. A later study indicated that forced-choice measures of personality exhibited useful incremental validity beyond cognitive ability alone, but not to a significantly different degree than did Likert-scored measures (Converse et al., 2008). Additionally, this study found that applicant reactions were less positive toward the forced-choice measure than they were toward the Likert-scored version (Converse et al., 2008).

Another limitation of forced-choice measures regards evidence that they may be influenced by cognitive ability. Christiansen et al. (2005) found that individuals with higher levels of cognitive ability were more successful at improving their scores on forced-choice inventories. They found that 6% of the variance in forced-choice Conscientiousness scores was explained by cognitive ability when participants were instructed to respond as job applicants, as compared to less than one-tenth of 1% of the variance explained in the condition with instructions to respond honestly (Christiansen et al., 2005). These results support the notion that responding in desirable ways to forced-choice measures is a cognitively demanding task, leaving this type of faking control effort more susceptible to individuals with higher levels of cognitive ability (Christiansen et al., 2005). Finally, it is also a concern that such formats force negative correlations between scales (Dilchert & Ones, 2011). Using the item above as an example, an individual stating that the item reflecting Conscientiousness is more like them is then necessarily not choosing the item reflecting Extraversion, although that individual may actually be high in both traits.

Methods that Attempt to Detect the Problem

Faking detection efforts have included the development of scales to identify faking (also known as validity scales, intentional-distortion scales, or social desirability scales), as well as item-response process models such as measuring the latency of response times and the examination of differential item functioning (Ellingson et al., 2007; Goffin & Christiansen, 2003; Kuncel & Borneman, 2007). I will begin with a discussion of validity scales, which have been suggested as appropriate to use in efforts to correct participants' scores on personality measures, or in the removal of participants whose scores are too extreme on such measures (Hough, 1998).

Probably two of the most recognized scales are the aforementioned L and K scales of the MMPI (Minnesota Multiphasic Personality Inventory). Originally, this test was designed to assess people suspected of mental health issues, although it has been used in personnel selection since the middle of the twentieth century (Butcher & Tellegen, 1966). The L scale is intended to detect blatant intentional dishonesty by identifying endorsements of attributes that are high in social respect but nearly impossible to meet (Framingham, 2011; Mesmer-Magnus & Visvesvaran, 2006). The K scale is intended to detect less overt faking attempts, where the image presented is overly positive but not impossible (Mesmer-Magnus & Visvesvaran, 2006).

Another fairly common measure of social desirability, developed for use with the Multidimensional Personality Questionnaire (MPQ), is Tellegen's unlikely virtues scale (Patrick, Curtin, & Tellegen, 2002). The MPQ is often used in selection contexts, and is closely related to both the NEO-PI and the MMPI (Tellegen & Waller, 2008). Similar to

the L and K scales of the MMPI, the unlikely virtues scale is comprised of items that represent qualities that are highly desirable but also improbable (Piedmont, McCrae, Riemann, & Angleitner, 2000). Other scales include Paulhus' (1984) Balanced Inventory of Desirable Responding (BIDR), which provides independent measures of self-deception and impression management. Paulhus (1984) differentiated between biased responding due to conscious impression management, and self-deception (which occurs when the respondent is unaware of the behavior) with the BIDR.

Myriad publications have addressed the effectiveness of such scales (Bagby, Buis, & Nicholson, 1995; Bagby, Gillis, & Dickens, 1990; Bagby, Rogers, Nicholson, Buis, Seeman, & Rector, 1997; Christiansen et al., 1994; Ellingson et al., 1999; Hough, 1998; Li & Bagger, 2006; Ones et al., 1996; Piedmont et al., 2000; Rosse et al., 1998; Viswesvaran & Ones, 1999; Zickar & Drasgow, 1996). Bagby et al. (1997) found that both clinical and non-clinical participants produced higher scores on the MMPI's L and K scales in a fake-good condition when compared to an honest condition. A meta-analysis of directed-faking studies conducted by Viswesvaran and Ones (1999) found that social desirability scores were around one standard deviation higher for faked versus honest conditions. Additionally, Rosse et al.'s (1998) study found that adjusting Conscientiousness scores according to scores on a measure of social desirability reduced the levels of response distortion in hired applicants for selection rates of less than 50%, although the effect was attenuated as selection rates increased. As previously discussed, some applicants high in response distortion would also be hired at various selection rates (Rosse et al., 1998). Relatedly, discrepant findings were also reported by Ellingson et al.

(1999), who found that corrected scores more accurately represented honest scores yet did not consistently produce more correct selection decisions.

Austin (1992) showed more mixed results. She found that the L scale was the best predictor of fake-good respondents (while also never producing false positive results), yet the K scale represented a poor indicator of fake-good respondents (Austin, 1992). Indeed, it has often been the case that validity scales do not evidence much efficacy in applied contexts. I have previously mentioned the results of Ones et al.'s (1996) meta-analysis that found that researchers were able to correct personality scores using measures of social desirability without affecting criterion-related validities. In another example, Piedmont et al. (2000) found a lack of utility for an array of validity indices across multiple samples.

Additionally, in a classic study Kroger and Turnbull (1975) found that participants could fake specific personality profiles on the MMPI without detection. Another study found that the impression management scale of the 16PF (one of the inventory's 16 non-cognitive scales) measured different underlying constructs for applicant (faking) versus non-applicant (honest) conditions (Cattell & Mead, 2008; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). Continuing, researchers have also noted the inability of such indices to distinguish between those with truly high levels of desired traits and those engaged in faking behavior (Griffith & Peterson, 2008; Kuncel & Borneman, 2007; McCrae & Costa, 1983). Finally, a study by Hurtz and Alliger (2002) found that an embedded unlikely virtues scale was also vulnerable to being coached against.

Paulhus' (1984) two-factor structure of the BIDR was developed using factor analysis and was supported by multiple studies, as reported in his early publication regarding the inventory. However, recent work from Li and Bagger (2006) call into question the usefulness of this inventory and its distinction between the two factors in applied contexts. Their meta-analysis found that: criterion-related validity was not attenuated when self-deception or impression management are corrected for in personality measures, that neither of the two factors predicted performance, and that the two factors were correlated with personality traits (Li & Bagger, 2006). They concluded that the practice of correcting scores on the basis of such validity scales is unwarranted (Li & Bagger, 2006). Additionally, Reeder and Ryan (2011) posit that the finding that various scales of social desirability load on one of the two factors of the BIDR suggests that these scales do not all measure the same construct, casting even more doubt upon their utility in applied settings.

Despite the mixed findings, potential for false positives, and vulnerability to coaching, it has been fairly common for researchers to correct inventories with high social desirability scale scores, or to simply remove such participants altogether (Goffin & Christiansen, 2003). However, these approaches, promulgated by Hough's (1998) paper proclaiming their reasonable effectiveness, may actually be ineffective or even detrimental. After finding that researchers were able to correct personality scores using measures of social desirability without affecting criterion-related validities, Ones et al. (1996) concluded that removing or correcting for social desirability might result in the loss of some true variance (i.e. predictive power) from the substantive part of the test. Ellingson et al. (2007) reached a similar conclusion. Noting that correlations between

scores (on intentional-distortion scales and personality traits such as Conscientiousness and Emotional Stability) confound faking with trait measurement, they questioned the appropriateness of using such scales as measures of applicant faking (Ellingson et al., 2007).

In fact, a multitude of empirical findings have suggested that these scales are correlated with substantive personality traits, leading researchers to question their use in excluding or correcting test scores (Ellingson et al., 2007; Li & Bagger, 2006; MacCann et al., 2011; Ones et al., 1996; Ones & Viswesvaran, 1998). Griffith and Peterson (2008) went so far as to suggest that such scales are poor representatives of faking and have no statistical relationship with the behavior. Uziel (2010) furthered this line of thinking with the suggestion that such scales should actually be redefined as identifying individuals that exercise high levels of self-control in social contexts. That article questioned the utility of such measures in validating self-report inventories and suggested that their real value lies in the measurement of substantive personality trait variance (Uziel, 2010).

Ones and Viswesvaran (1998) had earlier found as much with their meta-analytic results, reporting that social desirability may not be a good predictor of overall job performance, but that it does predict multiple variables that are important to work (such as job satisfaction, organizational commitment, and ratings of training success). A recent publication from Ispas, Iliescu, Ilie, Sulea, Askew, Rohlf, and Whalen (2014) further substantiates this notion. This study involving sales professionals found that impression management was not only associated with job performance, but that it also offered incremental validity over Conscientiousness and cognitive ability (Ispas et al., 2014).

The authors eventually concluded that terms such as faking and lie should no longer be used interchangeably with impression management scales, as they are inaccurate and serve to prevent potentially useful research into the utility of a construct that represents substantive trait variance (Ispas et al., 2014).

Indeed, the Ellingson et al. (1999) study (which involved Hough herself) reported earlier that corrections for social desirability often result in displacing honest individuals from top ranks, and concluded that such corrections are ineffective. Further, according to Goffin and Christiansen (2003), corrections treat social desirability as a suppressor variable, although social desirability has been shown not to be a useful suppressor variable for the criterion of job performance (Ones et al., 1996). Considering the mixed results, susceptibility to coaching, potential for removing valid trait variance, and the inability to distinguish between those with truly high levels of desirable traits and those engaged in intentional dissimulation, it seems that making score corrections using these scales should be avoided (Kuncel & Borneman, 2007).

From as early as 1983, publications reporting decreases in validity due to score corrections using validity scales can be found (McCrae & Costa, 1983). Noting the correlations of such scales with certain personality traits and the potential for confusing fakers with true honest high scoring respondents, McCrae and Costa (1983) questioned the practice of correcting scores using these scales. They may have said it best in their critique of social desirability scales when they noted that, “An individual who is in fact highly conscientious, well-adjusted, and cooperative would appear to be high in [social

desirability]. Paradoxically, it is the most honest and upstanding citizen that these scales would lead us to accuse of lying!” (McCrae & Costa, 1983, p. 883).

An alternative to correcting scores using these scales (or excluding individuals altogether) is to ask flagged individuals to take the test again. Ellingson, Heggstad, and Makarius (2012) reported that this approach resulted in more accurate scores (as opposed to initial scores) upon retesting when compared with a baseline measure. However, retested participants that had not engaged in intentional distortion were found to produce less accurate scores (Ellingson et al., 2012). While such results offer some promise for the much-maligned use of validity scales, the idea that this approach may undermine true trait representations with the potential for false-positives could prove its use difficult to justify.

Continuing with the discussion of faking detection efforts, the practices of measuring participants’ response latencies and examining differential item functioning (DIF) are next. These methods are based on attempts to understand or identify differences in the underlying response processes of diverse participants (Kuncel & Borneman, 2007; Mesmer-Magnus & Viswesvaran, 2006). The response latency approach is based on the notion that response time is influenced (positively or negatively) by the idiosyncratic systems of schemas held by the participants, and contends that when one chooses to respond to an item in a manner that is incongruent with one’s schematic system, the response will take longer (Holden & Hibbs, 1995). Such effects have been found in multiple studies.

For instance, Hsu, Santelli, and Hsu, (1989) reported that response latencies more

accurately indicated undergraduates engaged in directed-faking (in multiple contexts) than did scores from various validity scales. In another study, Popham and Holden (1990) reported finding larger latencies for participants that rejected relevant items, and smaller latencies for participants that endorsed relevant items. Later, Fekken and Holden (1992) reported that such response latency effects emerged for both directions (rejected and endorsed) of various schemas (positive and negative), even after standardization for individual- and item-level baselines.

While the results of such studies exhibit potential, there remains reason to question the applicability of this approach. For instance, it has been shown that participants are able to adopt a socially desirable schema for fake-good purposes, thereby decreasing response latencies (Holden, Kroner, Fekken, & Popham, 1992). Additionally, Vasilopoulos, Reilly, & Leaman (2000) found that job familiarity moderated the effect of impression management on response latency, such that as job familiarity increased the response latencies due to faking decreased. Relatedly, another study found that the ability of response latencies to detect faking is ineffective against fakers who have been coached to beat them (Robie, Curtin, Foster, Phillips IV, Zbylut, & Tetrick, 2000). Such possible confounds suggest that this method is far from a panacea, especially when coupled with the observation that the computer-based testing required for response latency measurement may not be possible, practical, or financially reasonable in most selection contexts (Kuncel & Borneman, 2007).

The analysis of DIF is a relatively new method with which researchers are attempting to detect faked responses. This line of analysis can take various forms and is

based on aspects of Item Response Theory (IRT), such as the idea that test items vary in their discrimination or difficulty between different populations, like those in honest contexts versus those in faking conditions (Levin & Zickar, 2002; Mesmer-Magnus & Viswesvaran, 2006). In other words, a particular item may have different response functions for different groups, such that otherwise similar individuals may differ in their probabilities for choosing a particular response option depending upon the context (Zickar & Robie, 1999). For example, one study by Stark et al. (2001) found that DIF occurred between applicant (faking) and non-applicant (honest) samples in each of the 15 examined non-cognitive scales of the 16PF. This study also reported that no single item type consistently evidenced DIF (Stark et al., 2001). Of particular interest to I/O Psychologists, another study found DIF between students and applicants for four of the six facets of Conscientiousness as measured by the NEO-PI-R (Griffin, Hesketh, & Grayson, 2004).

While the preceding promising results have emerged from this approach, along with others such as Zickar and Robie's (1999) report suggesting that participants found some items easier to fake than others, this method is not without its limitations. For instance, one group of researchers noted the interpretive difficulty associated with this method due to significant findings being highly dependent on sample size (Stark, Chernyshenko, & Drasgow, 2004). Stark et al. (2004) further noted that differential functioning did not manifest a decrease in the hiring of honest respondents, nor did it have much overall effect on the measurement characteristics of the 16PF. Therefore, the authors suggested that the (admittedly pervasive) statistically significant occurrence of DIF might have little practical significance (Stark et al., 2004). Still, researchers such as

Kuncel and Borneman (2007) persist in the belief that the auspices of this line of research warrant further investigation.

The Kuncel and Borneman (2007) Unusual Item Response Technique

Kuncel and Borneman (2007) reported some intriguing findings from their study that evaluated the notion that faking often results in complex response patterns for certain items. Although not based on IRT, their study is related to the DIF approaches previously mentioned in that it uses differential item response patterns in faking detection. The authors estimated that the complex patterns they found were due to participants holding disparate ideas as to which response option is maximally desirable (Kuncel & Borneman, 2007). Within this study, they identified items from Goldberg's (1992) 100 adjective markers (which were developed to represent the FFM) that evidenced unusual response distributions when comparisons were made between the directed-faking and honest conditions of a within-subjects design (Kuncel & Borneman, 2007). The authors then attempted to blindly identify whether the individual participants of a cross-validation sample were from the faking (or honest condition) by using a recoding scheme based on the response patterns of these unusual items (Kuncel & Borneman, 2007).

A critical concern of theirs was to avoid items that exhibited simple inflation of scores in faking contexts, which would be indistinguishable from truly high levels of desirable traits (Kuncel & Borneman, 2007). Having identified multiple items that fit their criteria, Kuncel and Borneman (2007) believed that summing across the items' recoded values for each participant would provide a faking indicator that would enable

the authors to accurately distinguish between faked responses and true endorsements of desirable responses in the cross-validation sample (Kuncel & Borneman, 2007). I will provide readers with a detailed discussion of their method in this section.

It is important to begin a discussion of this technique by elaborating upon the difference between the response distributions of typical items, and the response distributions that constituted what Kuncel and Borneman (2007) referred to as an unusual pattern. To aid in this discussion, Figure 1 reproduces two histograms from Kuncel & Borneman (2007) that represent the response distributions of a typical item. As one can see, the response options (labeled along the x-axes) ranged from one to nine. The y-axes indicate the number of participants whose responses for the adjective careful are represented above the respective response options (Goldberg, 1992; Kuncel & Borneman, 2007).

Figure 1a represents the honest condition, which evidenced a slight negative skew, with most people believing that they are above average for the adjective careful (Goldberg, 1992; Kuncel & Borneman, 2007). Figure 1b represents the faking condition, which evidenced a more extreme negative skew and higher overall endorsements for the adjective careful (Goldberg, 1992; Kuncel & Borneman, 2007). In a hiring situation, low scorers from either condition would likely not be selected, while at the high end it is impossible to differentiate between fakers and those who truly possess the desirable trait. This results in the responses lacking much utility for select-in purposes (Kuncel & Borneman, 2007).

Figure 1. A Typical Item's Response Distributions from Honest (a) and Faking (b) Conditions for the Test Item Careful (Goldberg, 1992; Kuncel & Borneman, 2007).

Figure 2 reproduces two additional histograms from Kuncel and Borneman's (2007) original publication that represent the response distributions of an unusual item. Here, the honest condition depicted in Figure 2a is only slightly skewed, with a clear central mode for the adjective *imperturbable* (Goldberg, 1992; Kuncel & Borneman, 2007). Figure 2b represents the faking condition, which is strikingly dissimilar. There appear to be three distinct modes, with high levels of endorsement for the adjective *imperturbable* at both extremes, as well as at the center response option (Goldberg, 1992; Kuncel & Borneman, 2007). A comparison of the two distributions allows for the identification of multiple response options that are unlikely to be endorsed by honest participants (Kuncel & Borneman, 2007).

Figure 2. An Unusual Item's Response Distributions from Honest (a) and Faking (b) Conditions for the Test Item Imperturbable (Goldberg, 1992; Kuncel & Borneman, 2007).

Having examined the paired response distributions (of both conditions) for each of the 100 Goldberg (1992) adjective markers, Kuncel and Borneman (2007) were able to identify 11 (10 tri-modal and one bi-modal) that fit their criteria for unusual items. For each of these 11 items, comprehensive comparisons of the frequency distributions of response option endorsements were made between the honest and faking conditions (Kuncel & Borneman, 2007). Using intervals of .5, the authors assigned faking indicator values ranging from -1 (low faking potential) to +1 (high faking potential) to every response option (for each item), with a neutral score of zero effectively representing a cut-score between faking and honest participants. Those response options that were endorsed more often in the faking condition received positive recoded values, while those

endorsed by a greater number of participants in the honest condition received negative recoded values.

Table 1 reproduces a one-item example from the original publication to aid in illustrating the manner in which this recoding scheme was established (Kuncel & Borneman, 2007). In Table 1, each response option (one through nine) for the sample item has both an honest and faking condition response frequency percentage (rounded to the nearest whole number) listed underneath. The authors judgmentally assigned the recoded value presented in the Scoring Key row depending upon whether the discrepancy between the listed frequencies for the respective conditions was determined to be large, moderate, or negligible (Kuncel & Borneman, 2007).

As Table 1 illustrates, the authors determined that response options one and nine for this item evidenced a large discrepancy (with more endorsements in the faking condition) and assigned these options recoded values of +1, while option eight evidenced a large discrepancy (with more endorsements in the honest condition) and received a recoded value of -1 (Kuncel & Borneman, 2007). Option two was deemed to have only a moderate discrepancy (with more endorsements in the faking condition) and received a recoded value of +.5, while options four, six, and seven were all deemed to have moderate discrepancies (with more endorsements in the honest condition) and were assigned recoded values of -.5. Options three and five evidenced equal percentages of endorsements across conditions, and received recoded values of 0 (Kuncel & Borneman, 2007).

Table 1. *Sample Recoding Scheme for One Item (Kuncel & Borneman, 2007).*

Response Option	1	2	3	4	5	6	7	8	9
Honest (%)	3	3	3	5	18	22	21	16	9
Faking (%)	15	6	3	2	18	15	11	7	24
Scoring Key	+ 1.0	+ .5	0	−.5	0	−.5	−.5	−1.0	+ 1.0

This process was repeated for all of the previously identified unusual items, resulting in a unique recoding scheme for each of those 11 items. All participants in the cross-validation sample were then assigned a recoded value (as dictated by this scheme) for each those 11 unusual items. Summing each participant's recoded values across all of the 11 unusual items resulted in what the authors regarded as a faking indicator for that individual (Kuncel & Borneman, 2007). Using zero as the cut-score, the authors then used these values to blindly predict whether participants from the cross-validation sample had been part of the faking condition with up to 78% accuracy, while producing a false positive rate of only 14%. Additionally, raising the cut score to minimize the false positives to a rate below 1% still allowed for the authors to detect faked tests at a rate as high as 37% (Kuncel & Borneman, 2007).

In addition to this method's apparent ability to accurately differentiate between those with truly high levels of desirable traits and those engaging in prevarication, Kuncel and Borneman (2007) noted many other benefits to their technique. They deemed it relatively coaching-resistant, as avoiding all extreme responses would result in low scores, whereas always endorsing them would often be viewed as an indicator of faking. They also reported that the method was not strongly correlated with any of the individual

difference measures implemented, which included: an additional personality test (MPQ), a social desirability scale (BIDR), and the Wonderlic (1992) measure of cognitive ability (Kuncel & Borneman, 2007).

While this method appears to address many of the common concerns regarding the potential for faking on personality measures, it is not without limitations. First, the study used college students (instructed to answer honestly at time one, and subsequently directed to fake on a second inventory) in a lab setting. Although using the within-subjects design allowed for analysis of faking at the individual-level and removed the possibility of sample characteristics causing differences between the two conditions (extant in between-subjects designs), the study is still limited by using a directed-faking technique which often serves to exaggerate differences between conditions (Mesmer-Magnus & Viswesvaran, 2006; Viswesvaran & Ones, 1999; Smith & Ellingson, 2002). As the belabored point in the literature maintains, one cannot be certain whether directed-faking in a lab setting is an accurate representation of faking in the real-world of personnel selection contexts, as the degree of faking may be increased and the variability between participants decreased due to this method (Abrahams et al., 1971; Hogan et al., 2007; Smith & Robie, 2004).

In addition, participants were directed “to imagine that they were applying for a desirable job” (Kuncel & Borneman, 2007, p. 226). The probability that hundreds of students imagined an array of diverse jobs may represent a problem with the internal validity of this study. Multiple studies have found that participants have the ability to form a priori hypotheses about the profiles of various jobs and to subsequently fake those

profiles with a degree of accuracy (Kroger & Turnbull, 1975; Raymark & Tafero, 2009). Such findings are reinforced by Birkeland et al.'s (2006) meta-analysis, which interpreted certain findings as suggesting that applicants distort their responses for personality dimensions that are viewed as job relevant. Extrapolating, rather than unusual response patterns being due to the nature of the item itself, they may have simply been due to differential views of the desirability of that item as it relates to the diverse occupations imagined by various students. Additional limitations of the previous study include: the authors' use of a qualitative, post hoc approach to develop the recoding scheme; the inclusion of Goldberg's (1992) adjective markers, which is rarely used in selection contexts and relies on single word items rather than the more typical statement presentation; as well as the reliance on an unusual nine-option response scale which deviates from more conventional five- or seven-option formats.

CHAPTER III

SUMMARY AND RESEARCH QUESTIONS

Summary

Although the degree of importance is still a topic of some contention, the susceptibility of personality measures to faking has been a continual concern of I/O Psychologists and has increased as the use of personality measures has continued to expand with modern selection practices. While many argue that faking does not represent a significant problem to the use of personality measures in hiring decisions, others have found that it can have a profound impact at the individual-level. This often occurs by displacing honest respondents from top positions when rank-ordering applicants, an effect which has repeatedly evidenced an inverse relationship with the size of selection rates. Offering incremental validity to the selection process and protecting honest responders from displacement are both important consequences that may result from addressing the potential problem of faking on personality measures. While sundry attempts have been made to develop a reliable method with which to address this issue, an acceptable method has evaded consensus up to this point.

The Kuncel and Borneman (2007) study offers a novel approach that evidenced encouraging results, while also possessing notable limitations. This study endeavored to

address several limitations of this approach to faking detection. First, this study will examine real-world applicants' scores on a personality measure as compared to their own previous scores on the same inventory (that was completed for research purposes 1 to 2 years prior). This type of within-subjects field study will allow for the assessment of individual change without relying on directed-faking in lab conditions, which is rare in faking research. Further, rather than using the method to predict from which condition (honest vs. directed-faking) the results of an inventory were obtained, this method provides a more accurate estimation of the effectiveness of the procedure by allowing for the identification of those indicated as high in faking potential that also evidenced score increases in a true application context.

In addition, the jobs applied for were all from the same family, which should serve to reduce variance in the responses of fakers due to hypothesizing disparate job profiles. Also, a quantitative, a priori recoding scheme was used to determine faking potential. This allows for the ease of replication, as well as reduces unnecessary bias or variance on the part of individual raters or due to differences in the judgment of distinct raters. Finally, this study used the NEO-PI-R, in place of Goldberg's (1992) adjective markers. The NEO-PI-R represents a well-validated and frequently used selection tool that incorporates a typical statement presentation of items and a more conventional five-option response format (Costa & McCrae, 1992).

Addressing these limitations offers further clarity as to the degree of practical utility of the Kuncel and Borneman (2007) approach. Once assessing its accuracy with these modifications, I examined its impact at various cut-scores in multiple select-in and

select-out contexts, with the goal of minimizing honest responder displacement and false positive faking identifications.

Research Questions

The following research questions were examined in the course of this study:

Research Questions 1A and 1B- Reflecting specific concerns set forth in Kuncel and Borneman (2007) regarding potential modifications to the method:

1A- Will this approach be functional when limited to only five response options?

1B- Will this approach break down because the stereotypes or schemas regarding the ideal candidate for one particular job family (and employed in faking efforts) are all relatively similar?

Research Question 2- Will this approach translate to real-world applicant research, as opposed to the directed-faking setting in which it was developed?

Research Question 3- Will making the aforementioned ameliorations impact the efficacy of the Kuncel and Borneman (2007) technique in identifying fakers at various cut-scores?

Research Question 4- Using Conscientiousness, Extraversion, and Neuroticism as predictors, what is the impact of multiple faking indicator cut-scores from this method on select-in decisions at various selection rates?

Research Question 5- Using Conscientiousness, Extraversion, and Neuroticism as predictors, what is the impact of multiple faking indicator cut-scores from this method on select-out decisions at various cut-offs?

CHAPTER IV

METHOD

Participants

For the current research, archival data was examined in an attempt to answer the research questions. Therefore, ethical concerns regarding research involving human subjects were largely minimized. Additionally, the dataset used contained no identifiers regarding the participants, so concerns over the protection of potentially sensitive information were not relevant.

The participants in this archival dataset were 213 Communications majors at a Romanian University, that later applied for various positions within the professional field of Communications. The participants ranged in age from 21 years to 37 years old ($M = 26.97$, $SD = 4.37$). The sample consisted of approximately equal numbers of men (110) and women (103).

Measures

The study used archival data that was previously collected from a sample of Communication majors of a Romanian university, who went on to be involved in various job application processes within the field of Communications. The data included the results of a personality inventory completed as part of the application process, as well as

the results of the same inventory previously administered for research purposes during the students' time in college. Regarding the typical concern over testing effects in within-subjects designs, this should not be an issue with this study as the respective inventories were completed several years apart (Mesmer-Magnus & Viswesvaran, 2006). The inventory completed was the Romanian version of the Revised NEO Personality Inventory (NEO-PI-R), which measures an individual on each of the five factors of the FFM (Costa & McCrae, 1992; Ispas et al., 2014). The NEO-PI-R is a 240-item personality measure that allows for a comprehensive assessment of normal adult personality, by including 30 eight-item scales that assess each of six of the most important facets that respectively define each of the five factors (Costa & McCrae, 1992). Item responses for the NEO-PI-R are made using a five-point Likert scale that ranges from zero (strongly disagree) to four (strongly agree).

The origins of the NEO-PI-R can be traced back to Costa and McCrae's 1978 NEO Inventory, which measured facets under the factors of Neuroticism, Extraversion, and Openness to Experience (Costa & McCrae, 1997). Adding global scales for Conscientiousness and Agreeableness in 1985, Costa and McCrae republished the inventory as the NEO-PI (Costa & McCrae, 1997). The NEO-PI-R is Costa and McCrae's (1992) revision to the NEO-PI that effectively culminated over 15 years of research. This revision offers improvements to several original items that allow for more measurement accuracy and includes the addition of facet scales for Agreeableness and Conscientiousness (Costa, 1996; Costa & McCrae, 1992). There is also a short (60 item) version of the NEO-PI-R that is referred to as the NEO-FFI and is scored at the factor level only (Costa & McCrae, 1992). The widespread acceptance of the Costa and

McCrae's work prompted Salgado (1997) to note that their labels for the five factors are generally the most accepted, although he did acknowledge that the factor labels vary among researchers to some degree. This is evidenced by the fairly common use of Emotional Stability (as interchangeable with reverse-scored Neuroticism) that can be witnessed in many publications (Barrick & Mount, 1991; Hills & Argyle, 2001; Hogan & Holland, 2003; Salgado, 1997; Ziegler et al., 2011).

NEO-PI-R sample items for each of the five factors include: for Neuroticism, "I am not a worrier;" for Extraversion, "I sometimes fail to assert myself as much as I should;" for Agreeableness, "I would hate to be thought of as a hypocrite;" for Conscientiousness, "When a project gets too difficult I decline and start a new one;" and for Openness, "I think it's interesting to learn and develop new hobbies" (Costa & McCrae, 1992, pp. 68-74). The factors of Neuroticism (or Emotional Stability, reverse-scored), Extraversion, and Conscientiousness will be examined in this study (Costa & McCrae, 1992; Hills & Argyle, 2001; Ziegler et al., 2011).

Sample items for each facet under Neuroticism include: for Anxiety, "I am easily frightened;" for Angry Hostility, "I am known as hot-blooded and quick-tempered;" for Depression, "Sometimes I feel completely worthless;" for Self-Consciousness, "At times I have been so ashamed I just wanted to hide;" for Impulsiveness, "I have trouble resisting my cravings;" and for Vulnerability, "It's often hard for me to make up my mind" (Costa & McCrae, 1992, pp. 68-69). Sample items for each facet under Conscientiousness include: for Competence, "I'm known for my prudence and common sense;" for Order, "I keep my belongings neat and clean;" for Dutifulness, "I pay my

debts promptly and in full;” for Achievement Striving, “I work hard to accomplish my goals;” for Self-discipline, “Once I start a project, I almost always finish it;” and for Deliberation, “I think things through before coming to a decision” (Costa & McCrae, 1992, pp. 73-74).

Since its development, the NEO-PI has been widely used in I/O Psychology for studies regarding the predictive ability of personality, selection, and faking (Costa, 1996; Denis, Morin, & Guindon, 2010; Furnham, 1997; Piedmont & Weinstein, 1994; Winkelspecht et al., 2006). In addition, the test’s developers (Costa and McCrae) have participated in multiple publications chronicling its validity, reliability, utility, and generalizability (Costa, 1996; Costa & McCrae, 1992; Costa & McCrae, 1997; McCrae & Costa, 1987; McCrae & Costa, 1997; McCrae, Costa, Del Pilar, Rolland, & Parker, 1998). I will begin my discussion of such reports with a review of some of the publications involving the authors of the inventory. I will then proceed into a review of some additional publications that report findings involving the NEO-PI-R as it relates to I/O Psychology.

To begin, the professional manual that accompanies the NEO-PI-R provides extensive data chronicling the use and characteristics of the inventory. Regarding internal consistency, coefficient alphas for the five factors range from .87 to .92, with Neuroticism (.92) and Conscientiousness (.90) being the two highest (Costa & McCrae, 1992). Coefficient alphas for the individual facets under Neuroticism range from .68 to .81, while those under Conscientiousness range from .62 to .75 (Costa & McCrae, 1992). Multiple studies regarding the (short-term and long-term) test-retest reliability of versions

of the inventory are also reported in the manual. In a three-month lapse between assessments of the NEO-FFI and the NEO-PI-R, college students evidenced coefficients of .79 for Neuroticism and .83 for Conscientiousness (Costa & McCrae, 1992). A three-year study reported a coefficient of .79 for Conscientiousness as scored by the NEO-PI, and a six-year study reported coefficients ranging from .68 to .83 (in both self-reports and spouse ratings) for Neuroticism, Extraversion, and Openness as scored by the NEO-PI (Costa & McCrae, 1992).

The professional manual also reports on the construct validity of the inventory as supported by multiple studies, including: substantial correlations between NEO-PI factors and Goldberg's (1992) adjective markers for the FFM, and correlations between the NEO-PI and the Hogan Personality Inventory (HPI) that is also based on the FFM (Costa & McCrae, 1992; Hogan & Hogan, 1989). In addition, the authors report support for convergent validity as evidenced by correlations between similar constructs on the NEO-PI-R and alternative self-report measures, as well as by the agreement between self-reports and observer ratings (Costa & McCrae, 1992). The authors also report support for discriminant validity as evidenced by the negative relations between dissimilar constructs on the NEO-PI-R and similar measures, and by near-zero correlations between self-reports and observer ratings between factors (Costa & McCrae, 1992). Continuing, in a cross-cultural study assessing the generalizability of the NEO-PI-R and its recent translation to multiple languages, McCrae et al. (1998) reported many similarities between the United States and other cultures.

Of particular relevance to the current research is the generalizability of the NEO-PI-R to Romanian samples. Ispas, Iliescu, Ilie, and Johnson (2014) found considerable evidence suggesting that the Romanian translation of the NEO-PI-R has similar psychometric properties when compared with normative samples (Ispas et al., 2014). The authors' use of factor analysis revealed a factor structure for the NEO-PI-R in a large Romanian sample that was similar to that found in American samples (Ispas et al., 2014). Also, internal consistencies and test-retest reliabilities were found to be similar to those from other translated versions of the test (Ispas et al., 2014). Furthermore, convergent, discriminant, and construct validity were also evidenced through the use of self-other agreement, as well as through comparisons with similar measures of the FFM (Ispas et al., 2014). In particular, Conscientiousness was found to have a coefficient alpha of .90 (with those of the individual facets ranging from .64 to .72), test-retest reliability of .73, and self-other agreement of .50 (Ispas et al., 2014). Neuroticism was found to have a coefficient alpha of .91 (with those of the individual facets ranging from .68 to .77), test-retest reliability of .79, and self-other agreement of .46 (Ispas et al., 2014). These figures all bear remarkable similarity to corresponding figures reported by McCrae and Costa (1992) in the test's professional manual.

The NEO-PI-R has evidenced utility specific to work contexts as well, with Neuroticism and Conscientiousness often exhibiting primary importance. Costa (1996) published a compilation of research findings regarding the application of the NEO-PI-R in I/O Psychology. In this article, he related earlier findings from Costa, McCrae, and Holland (1984), which reported that Extraversion, Agreeableness, and Openness were

related to vocational interests. In a subsequent replication focused only on Openness, he reported, similar results were found (Costa, 1996; Holland, Johnston, Hughey, & Asama, 1991). Offering some criterion-related validity, Costa (1996) cited findings from Piedmont and Weinstein's (1994) study that reported correlations between corresponding facet scales (under Neuroticism and Conscientiousness, as well as under Extraversion and Agreeableness) of the NEO-PI-R and supervisory ratings.

Continuing, Costa, McCrae, and Kay (1995) found that candidates recommended for hire as police officers (by trained psychologists) also scored higher on all six Conscientiousness facets and lower on all six Neuroticism facets of the NEO-PI-R. Summarizing findings reported by Gandy, Dye, and MacLane (1994), Costa (1996) notes that the strongest significant correlations between the NEO-PI-R and supervisory ratings (in both men and women) were found for Conscientiousness. These relations were maintained even after controlling for age and education (Costa, 1996). Finally, in a recent study using the French translation of the NEO-PI-R, Denis et al. (2010) reported that a facet of Conscientiousness predicted supervisory ratings of task performance, while facets under Neuroticism predicted supervisory ratings of both task performance and contextual performance in a French-Canadian sample.

Relevant to this study, Iliescu, Ilie, Ispas, and Ion (2012) reported correlations between the factors of the FFM (as measured by the Romanian NEO-PI-R) and subjective (customer orientation and persuasion, other-ratings), objective (financial indicators, attainment of objectives), and overall job performance for multiple professions. Neuroticism evidenced correlations of -.15, -.20, and -.20 respectively with

measures of objective, subjective, and overall job performance for public servants and -.12 for overall performance of public hospital CEO's (Iliescu et al., 2012).

Conscientiousness evidenced correlations of .24, .26, and .31 respectively with measures of objective, subjective, and overall job performance for public servants and .28 for overall performance of public hospital CEO's (Iliescu et al., 2012). In a subsequent study that involved a representative sample of Romanian nationals and also used the Romanian NEO-PI-R, Iliescu, Ilie, Ispas, and Ion (2013) reported correlations of -.06 and -.24 respectively between Neuroticism and supervisor or patient ratings of job performance. This study also reported correlations of .32 and .22 respectively between Conscientiousness and supervisor or patient ratings of job performance (Iliescu et al., 2013).

Procedure

To answer the research questions listed above, I began by following the approach set forth by Kuncel and Borneman (2007), and explained in the section above that describes their technique. First, I compared the histograms for each NEO-PI-R item between the two conditions (research vs. applicant), and identified any items that evidenced the unusual pattern described above.

Research Questions 1A and 1B

This initial phase enabled me to analyze some of my preliminary research questions, regarding whether the unusual item response technique is functional when limited to only five response options and whether the approach breaks down when dealing with candidates from one particular job family.

No NEO-PI-R items were found to evidence the change from a somewhat normal distribution to the multimodal distribution type referenced in Kuncel and Borneman (2007). However, changes were found from the research context to the applicant context that still fit Kuncel and Borneman's (2007) main criteria for indicating faking behavior. These changes typically took one of two forms. The first form involved a distribution with low levels of extreme endorsements (response options 0 and 4) in the research context evidencing substantial increases in endorsements for both extreme response options in the applicant condition. This indicates not only changing responses on the part of the applicants, but also some disagreement as to which option would be viewed as most desirable by the organization. Figure 3, which displays the respective endorsement levels between conditions for test item 123 (representing the fantasy facet of Openness), provides an example of such an item. Figure 3a (research condition) shows fairly low (both below 10%) endorsement levels for options 0 and 4, and fairly high levels (all around 30%) for the other options. In Figure 3b (applicant condition) endorsements for both extreme response options more than doubled.

Figure 3. An Unusual Item's Response Distributions from Research (a) and Applicant (b) Conditions for Item 123 Representing the Fantasy Facet of Openness (Costa & McCrae, 1992).

The second form of change involved a skewed distribution in the research context transforming into a more normal distribution. This generally involved high levels of endorsements for two of the middle three response options (options 1, 2, and 3) and low levels of endorsement for the third in the research condition. In the applicant condition, the middle response option with the low levels of endorsements showed a drastic increase in endorsements, while the other two middle options remained relatively highly endorsed as well, although they necessarily decreased to some degree. Again, this indicates not only changing responses on the part of the applicants, but also some disagreement as to which response options offer maximal desirability. Figure 4, which displays the respective endorsement levels between conditions for test item 21 (representing the impulsiveness facet of Neuroticism), provides an example of this second type of item. Figure 4a (research condition) shows high levels of endorsements for response options 1 and 2 and much lower levels for option 3. In Figure 4b (applicant condition) the

endorsements for option 3 have increased substantially, although options 1 and 2 are still endorsed at relatively high levels.

Figure 4. An Unusual Item's Response Distributions from Research (a) and Applicant (b) Conditions for Item 21 Representing the Impulsiveness Facet of Neuroticism (Costa & McCrae, 1992).

In total, I found that over 17% (42/240) of the test items resulted in unusual distributions between contexts. Five of these items represented Neuroticism, eight represented Extraversion, 19 represented Openness, six represented Agreeableness, and four represented Conscientiousness.

Exploratory Inter-rater Agreement

Post hoc inter-rater agreement analyses were conducted for all NEO-PI-R items as an exploratory measure. Although these analyses were not involved in determining the final set of items used in calculating the faking indicators, the results may offer useful information toward future research regarding item selection, as well as a method of

quantifying this process necessarily relies heavily on qualitative judgment. For this exploratory procedure, a panel of four raters (graduate students in either I/O or Quantitative Psychology from a large Midwestern university, with knowledge of the current study) was established. This panel was tasked with assigning a rating (on a Likert style scale, ranging from one to seven) to each item, representing that item's relative strength or weakness as an indicator of faking behavior. A rating of seven indicated the best potential as a faking indicator, an item rated as a one showed the least potential, and those rated as fours were undetermined or neutral.

To begin this process, each rater received a set of instructions outlining the difference between typical and unusual items, which also highlighted the essential criteria (changing of scores and disagreement amongst participants) for an item's set of response-option distributions to qualify as unusual. The instructions also included one example each of the two forms of unusual items that had been identified through the initial item-selection procedure. These instructions were accompanied by histograms that depicted the response-option distributions (by percentage of participants) for all 240 NEO-PI-R items, from both the research and applicant conditions. One item at a time, the raters compared the research and applicant response-option histograms and assigned their faking indicator ratings in a process that took most several hours to complete.

Once the ratings for all 240 NEO-PI-R items were received from all four raters, inter-rater agreement (calculated with r_{wg}) was established respectively for each individual NEO-PI-R item, and collectively for all 240 items and for the 42 items selected for use in the respective faking indicator recoding schemes. The r_{wg} index is a

measure of inter-rater agreement that assesses the degree of consensus among raters, and is typically used in determining the appropriateness of combining data for higher-level analysis (Castro, 2002). The significance of the r_{wg} index has commonly been assessed at a criterion of .70, such that variables with indexes above that level have been deemed to have a high degree of consensus among raters (Castro, 2002).

Following the exposition set forth in James, Demaree, and Wolf (1984), r_{wg} for a single item was calculated by subtracting from one the quantity of the observed variance of item judgments multiplied by the expected variance if all judgments were due exclusively to random error. In the formula, $r_{wg(1)} = 1 - (s_x^2 / \sigma_{EU}^2)$, s_x^2 is the observed variance of the item and σ_{EU}^2 is the variance that would be expected if all judgments were due exclusively to random error. The second term (σ_{EU}^2) is calculated by subtracting one from the squared number of response options in the scale and dividing the resulting quantity by 12. In the formula, $\sigma_{EU}^2 = (A^2 - 1) / 12$, A corresponds to the number of response options in the rating scale (in this case seven). Additionally, as per recommendations set forth in James, Demaree, and Wolf (1984), items with an s_x^2 that exceeded the σ_{EU}^2 were recoded as $r_{wg(1)} = .00$.

Also following James, Demaree, and Wolf's (1984) formula, r_{wg} for multiple items was calculated as $r_{wg(J)} = J [1 - (s_x^2 / \sigma_{EU}^2)] / J [1 - (s_x^2 / \sigma_{EU}^2)] + (s_x^2 / \sigma_{EU}^2)$. In this formula, J corresponds to the number of parallel items for which inter-rater reliability is currently being assessed and s_x^2 becomes the mean of the observed variances for those J items (σ_{EU}^2 represents the same value as in the previous formula). For all 240 NEO-PI-R

items collectively $r_{wg(J)} = .99$, while for the 42 items selected for use in recoding collectively $r_{wg(J)} = .97$.

The means, standard deviations, skewness, kurtosis, and range (for both honest and faking conditions) for each of the unusual items were analyzed. Regarding the unusual items selected, all but one evidenced a range that included endorsements for all response options. Additionally, the direction of skewness per item tended to remain stable from the research context to the applicant context, and no items evidenced an extreme skewness that exceeded 1.0 (with cases in which the sign changed generally evidencing one of the two contexts remaining close to neutral). Kurtosis statistics were generally negative (with only a few exceptions, all of which were found in the research condition), indicating that most endorsements did not fall at the extreme response options. These statistics, along with the single-item $r_{wg(I)}$ scores, paired-samples t-statistics, and effect sizes (Cohen's d), are presented in Table 2.

Table 2. *Descriptive Statistics for the 42 Unusual Items, with Contrasts from the Research Condition to the Applicant Condition.*

#	Fct	<i>M</i>	<i>SD</i>	Skew	Kurt	Lo	Hi	<i>r</i> _{wg}	<i>t</i>	<i>p</i>	<i>d</i>
3	O	2.62	0.82	-0.62	0.30	0	4	.00	-4.58	.00**	-0.31
		2.37	0.93	-0.12	-0.52	0	4				
7	E	2.33	1.18	-0.14	-1.17	0	4	.00	6.59	.00**	0.45
		2.77	1.00	-0.42	-0.78	0	4				
21	N	1.66	0.97	0.30	-0.34	0	4	.00	5.03	.00**	0.35
		2.02	0.97	-0.04	-0.47	0	4				
37	E	2.68	1.00	-0.64	-0.11	0	4	.45	-6.82	.00**	-0.47
		2.26	1.03	-0.26	-0.52	0	4				
47	E	2.74	0.85	-0.60	0.33	0	4	.00	-12.82	.00**	-0.88
		1.93	0.91	-0.13	-0.12	0	4				
49	A	1.26	0.99	0.97	0.66	0	4	.33	7.79	.00**	0.53
		1.77	0.85	0.04	-0.27	0	4				
52	E	2.38	1.13	-0.45	-0.58	0	4	.31	-6.85	.00**	-0.47
		1.89	0.97	-0.08	-0.64	0	4				
60	C	2.80	0.99	-0.92	0.51	0	4	.88	-12.64	.00**	-0.87
		1.97	0.94	0.03	-0.20	0	4				
61	N	1.67	0.94	0.40	-0.70	0	4	.63	5.22	.00**	0.36
		1.99	1.04	0.08	-0.62	0	4				
71	N	1.51	0.94	0.50	-0.45	0	4	.70	7.96	.00**	0.55
		2.04	1.01	0.09	-0.48	0	4				
78	O	1.26	0.76	0.78	1.21	0	4	.83	13.73	.00**	0.94
		2.08	0.89	0.29	-0.30	0	4				
81	N	1.76	0.95	0.31	-0.38	0	4	.00	5.74	.00**	0.39
		2.12	0.97	0.10	-0.44	0	4				
93	O	1.77	0.98	0.41	-0.56	0	4	.00	1.29	.20	0.09
		1.85	0.93	-0.02	-0.36	0	4				
94	A	2.39	0.90	-0.36	-0.80	0	4	.00	-7.67	.00**	-0.53
		1.88	1.00	0.07	-0.54	0	4				
97	E	2.66	1.03	-0.60	-0.21	0	4	.95	-10.08	.00**	-0.69
		1.92	0.96	0.02	-0.43	0	4				
118	O	2.52	0.84	-0.41	-0.09	0	4	.75	-7.58	.00**	-0.52
		2.00	0.91	-0.18	-0.35	0	4				

#	Fct	<i>M</i>	<i>SD</i>	Skew	Kurt	Lo	Hi	<i>r</i> _{wg}	<i>t</i>	<i>p</i>	<i>d</i>
120	C	2.54	1.06	-0.49	-0.67	0	4	.70	-9.30	.00**	-0.64
		2.00	1.02	-0.12	-0.61	0	4				
123	O	2.12	1.01	-0.07	-0.76	0	4	.25	-1.21	.23	-0.08
		2.04	1.28	-0.05	-1.06	0	4				
136	N	1.61	1.02	0.50	-0.33	0	4	.94	3.68	.00**	0.25
		1.85	0.97	0.15	-0.59	0	4				
138	O	1.54	0.87	0.62	0.07	0	4	.13	0.65	.57	0.04
		1.58	1.19	0.39	-0.71	0	4				
153	O	1.68	0.85	0.34	0.08	0	4	.25	0.66	.51	0.05
		1.72	1.22	0.17	-0.95	0	4				
154	A	2.46	0.94	-0.49	-0.22	0	4	.45	-6.61	.00**	-0.45
		2.08	0.99	-0.14	-0.69	0	4				
158	O	2.39	1.00	-0.22	-0.48	0	4	.58	-0.50	.62	-0.03
		2.36	1.24	-0.31	-0.92	0	4				
163	O	2.46	0.94	-0.34	-0.51	0	4	.95	-1.29	.20	-0.09
		2.38	1.24	-0.38	-0.82	0	4				
168	O	2.18	0.92	-0.25	-0.92	0	4	.94	-0.45	.65	-0.03
		2.15	1.28	-0.13	-1.00	0	4				
173	O	2.19	1.10	-0.17	-0.89	0	4	.81	-0.57	.57	-0.04
		2.15	1.28	-0.13	-1.03	0	4				
177	E	2.79	0.79	-0.37	0.16	0	4	.44	-10.37	.00**	-0.71
		2.15	0.93	-0.30	-0.27	0	4				
180	C	2.44	1.02	-0.34	-0.42	0	4	.88	-8.27	.00**	-0.57
		1.95	0.97	0.06	-0.31	0	4				
183	O	2.30	0.97	-0.13	-0.84	0	4	.83	-2.68	.01**	-0.18
		2.13	1.22	-0.09	-0.87	0	4				
193	O	2.54	0.91	-0.29	-0.56	0	4	.00	-1.42	.16	-0.10
		2.46	1.13	-0.33	-0.59	0	4				
198	O	2.15	0.98	-0.13	-0.59	0	4	.95	-0.68	.50	-0.05
		2.11	1.36	-0.11	-1.20	0	4				
202	E	1.73	0.97	0.32	-0.52	0	4	.63	7.24	.00**	0.50
		2.22	0.93	0.07	-0.45	0	4				
209	A	2.56	0.95	-0.76	0.25	0	4	.58	-3.68	.00**	-0.25
		2.31	0.94	0.06	-0.51	0	4				
213	O	1.98	1.01	-0.01	-0.93	0	4	.58	-0.74	.46	-0.05
		1.93	1.26	0.09	-0.96	0	4				

#	Fct	<i>M</i>	<i>SD</i>	Skew	Kurt	Lo	Hi	<i>r_{wg}</i>	<i>t</i>	<i>p</i>	<i>d</i>
218	O	1.82	0.99	0.16	-0.67	0	4	.25	0.16	.87	0.01
		1.83	1.24	0.14	-1.00	0	4				
220	C	2.40	1.14	-0.56	-0.54	0	4	.83	-5.28	.00**	-0.36
		2.05	1.00	-0.07	-0.55	0	4				
223	O	2.38	1.00	-0.23	-0.75	0	4	.69	-1.82	.07	-0.12
		2.27	1.28	-0.13	-1.06	0	4				
228	O	1.93	0.88	0.01	-0.90	0	4	.00	0.35	.73	0.02
		1.95	1.17	0.07	-0.84	0	4				
229	A	2.54	1.00	-0.61	-0.26	0	4	.83	-7.87	.00**	-0.54
		2.07	0.97	-0.07	-0.50	0	4				
233	O	2.59	0.79	-0.37	-0.26	1	4	.45	-1.69	.09	-0.12
		2.49	1.07	-0.35	-0.53	0	4				
237	E	2.54	0.92	-0.41	-0.41	0	4	.81	-6.45	.00**	-0.44
		2.12	0.89	0.14	-0.28	0	4				
239	A	1.64	0.92	0.85	0.17	0	4	.83	4.01	.00**	0.27
		1.86	0.96	0.16	-0.58	0	4				

Note. The split cells for *M*, *σ*, skewness, kurtosis, and range are divided such that the statistic for the research condition is presented above the line and that for the applicant condition is presented below the line. ** denotes $p < .01$. *M* represents the mean response option endorsement for the item. *SD* represents the standard deviation for the sample's endorsements per item. Low and High represent the range of scores from lowest to highest response option endorsed. *r_{wg}* represents the interrater agreement for each item's potential as a faking indicator. *t* represents the test statistic for the difference in means between the two conditions for each item, and *p* represents the significance level (probability of the difference being due to chance) of that statistic. Positive values for *d* (the magnitude of the effect, uninfluenced by sample size) represent increases (from the research to the applicant condition) in the mean response option endorsements for that item.

Having identified the items that I felt best fit the criteria, I then recoded the set of response-options for each item. However, unlike in the Kuncel and Borneman (2007) study, this was done using proportions of the respective percentages per condition for

each response, rather than qualitative judgment as to the degree of discrepancy between them. The smaller percentage of endorsers for each item response option was divided by the larger percentage of endorsers, which resulted in a ratio that represents the relative proportion of respondents from the less-represented condition of that response option, as compared to respondents from the alternative condition. If the research condition was more-represented, then the recoded value was assigned a negative value to signify lower levels of faking potential; if the applicant condition was more-represented, then the recoded value was assigned a positive value to signify higher levels of faking potential.

The recoding values were based on Cohen's (1988) recommendations for describing effect sizes as small (.2), medium (.5), and large (.8). However, as smaller proportions actually represented larger discrepancies here, the inverse was the case. This resulted in a recoding scheme in which values $\leq .2$ were deemed large, those from $> .2$ to $\leq .5$ were deemed medium, those from $> .5$ to $\leq .8$ were deemed small, and those from $> .8$ to ≤ 1 were deemed equivalent. The large ratios were then assigned values of ± 3 , the medium ratios were assigned values of ± 2 , the small ratios were assigned values of ± 1 , and the equivalent ratios were assigned a value of 0. For test item 21 referenced above, this scheme resulted in the following recoding scheme: option 0 = $5.2/9.9 = .53$ = small (non-faking) = -1, option 1 = $24.9/37.1 = .67$ = small (non-faking) = -1, option 2 = $33.8/38 = .89$ = equivalent = 0, option 3 = $16/26.8 = .60$ = small (faking) = +1, and option 4 = $3.3/5.2 = .63$ = small (faking) = +1. The recoding scheme for this item is presented in Table 3.

Table 3. *Sample Recoding Scheme for Item 21 Representing the Impulsiveness Facet of Neuroticism.*

Response Option	0	1	2	3	4
Research %	9.9	37.1	33.8	16.0	3.3
Applicant %	5.2	24.9	38.0	26.8	5.2
Recoded Value	-1 (-.5)	-1 (-1)	0 (+.5)	+1 (+1)	+1 (+.5)

Note. Estimated recoded values from an attempt to recreate Kuncel and Borneman's (2007) qualitative method are listed in parentheses.

Establishing this quantitative recoding scheme a priori was expected to offer many advantages to the judgmental, post hoc approach used by Kuncel and Borneman (2007). In addition to eliminating any variance and/or bias due to rater judgment (thereby facilitating replication), recoding in this way ensures that the overall context of responses is represented. For instance, a difference of five between conditions becomes more meaningful when found between 5% and 10% of responders (which would result in a medium ratio of .5 and a recode value of +/-2) than it is when between 25% and 30% of responders (which would result in an equivalent ratio of .83 and a recode value of 0). Alternatively, with the Kuncel and Borneman (2007) qualitative approach it is likely that both of these differences between conditions would have been deemed small and recoded with the same value (+/- .5), even though one represents a doubling of the percentage of endorsers while the other represents what is seemingly a negligible difference. Further, this method results in an additional recoded value both above and below 0 (providing more precision to the scoring scheme), which should increase the identifying efficacy of the technique. Due to these implications, I believe this method represents a significant improvement to the original design. To analyze the difference between my refinements

and the original method, I also attempted to recreate Kuncel and Borneman's (2007) judgmental method for comparative purposes.

Once each of the unusual items was recoded, the recoding schemes for each respective unusual item were then used to rescore each completed inventory from the application condition. The resulting values were then summed for each individual inventory to produce a faking indicator for that individual. In addition, true-faking categorizations (whether or not the individual actually faked on in the applicant condition) were assigned using multiple methods.

Initially, several methods for determining which participants were faking were examined respectively for both predictors. These methods included: Standard Error of Measurement (SEM) with 95% confidence intervals built around the honest score alone and around both scores; Standard Error of Difference (SED) with 95% confidence intervals built around the honest score alone and around both scores; a 95% confidence interval built around change scores using the Standard Error of Measurement for the Difference Score (SEM_d); an attempt to use reliability of the change scores to calculate SEM of the change scores; whether or not a participant's change score exceeded a threshold of 1 SD beyond the mean change score (regardless of direction) of the entire sample; and examining whether a participant's change score from the honest condition to the applicant condition exceeded a threshold of $\frac{1}{2}$ SD (honest condition) in either direction (Arthur, Glaze, Villado, & Taylor, 2010; Griffith et al., 2007; Hogan et al., 2007; McFarland & Ryan, 2000). Appendix A contains a detailed discussion of each of these methods. Table 4 presents the findings from this preliminary examination.

Table 4. *Preliminary Findings Regarding Applicability of Various Methods for Categorizing True Fakers.*

Categorization Method	Predictor		
	Conscientiousness	Neuroticism	Extraversion
SEM (1 CI)	11/213	10/213	1/213
SEM (2 CI)	0/213	0/213	0/213
SED (1 CI)	4/213	1/213	1/213
SED (2 CI)	0/213	0/213	0/213
SEM _d	146/213	114/213	99/213
SEM (a Change)	N/A	N/A	N/A
>+/- 1SD+ M Change	28/213	33/213	53/213
> +/- ½ SD Change	67/213	42/213	53/213

Note. Findings are presented as the number of participants categorized as faking out of the total in the sample. Change score reliabilities were found to be negative with this dataset, and were therefore unusable.

Considering this data, it becomes clear that only three of the methods examined yielded a sufficient number of true faking categorizations to examine the detection method in question. Further, given that well over half of the sample (for one of the respective predictors) was regarded as a faker with the SEM_d approach, it was concluded that this method of categorization was too lenient toward faking conclusions. Similarly, considering that so few categorizations were made with the SEM (1 and 2 CI) and SED (1 and 2 CI) methods, it was concluded that these approaches were too conservative against faking conclusions. Therefore, the > +/- ½ SD Change and > +/- 1SD + |M

Change| methods were used to categorize true fakers for this study.

As discussed above, the $> \pm 1SD + |M \text{ Change}|$ method (subsequently referred to as $\frac{1}{2} SD$) relied upon the mean difference (M_D) between research condition scores and application condition scores for Conscientiousness ($M = 6.41$, $SD = 7.95$), Neuroticism ($M = -3.35$, $SD = 7.87$), and Extraversion ($M = 2.25$, $SD = 7.44$). The absolute value of the sum of the SD of the difference scores and the M_D , resulted in a threshold of ± 14.43 for change in Conscientiousness scores, ± 11.22 for Neuroticism scores, and ± 9.69 for Extraversion scores. Change in either direction beyond these respective thresholds resulted in a true faking categorization. For Conscientiousness, approximately 13% (28/213) of the sample was found to have exceeded this limit with their change in scores and were subsequently labeled true fakers. For Neuroticism, approximately 15% (33/213) of the sample was found to have either raised or lowered their scores beyond this limit. For Extraversion, approximately 25% (53/213) of the sample was found to have either raised or lowered their scores beyond this limit.

The $> \pm \frac{1}{2} SD$ Change method (subsequently referred to as $\frac{1}{2} SD$) used thresholds determined by the observed SD from the honest condition. If participants changed their scores in the faking condition by more than $\frac{1}{2} SD$ (honest condition), then those participants were labeled as fakers. For Conscientiousness ($SD = 20.15$), this resulted in a threshold of ± 10.07 with approximately 31% (67/213) of the sample found to have either raised or lowered their scores beyond this limit and subsequently labeled true fakers. For Neuroticism ($SD = 20.83$), this resulted in a threshold of ± 10.42 with approximately 20% (42/213) of the sample found to have either raised or lowered their

scores beyond this limit. For Extraversion ($SD = 18.40$), this resulted in a threshold of ± 9.20 with approximately 25% (53/213) of the sample found to have either raised or lowered their scores beyond this limit. Of note here is that the respective thresholds for Extraversion (1 $SD = \pm 9.69$ and $\frac{1}{2} SD = \pm 9.20$) resulted in the same decisions, as a score change of 10 or greater (as score changes always occurred in the form of whole numbers) was required for both methods to result in a faking categorization.

The faking indicator scores for each predictor were referenced against the true faking categorizations (determined using the respective 1 SD and $\frac{1}{2} SD$ methods) for each participant to determine the potential of the Kuncel and Borneman (2007) method to identify faking at various cut-scores ($\geq 0, 1$, and 2 standard deviations above the mean faking indicator score). Inventories with indicator scores above the cut-score were expected to belong to individuals identified as fakers (as defined by application scores outside of the previously mentioned extreme limits of the respective confidence intervals) in the application context, while those below the cut-score were expected to belong to individuals not identified fakers (similarly defined as application scores within or below the extreme limit of the respective confidence intervals). Additionally, as the cut-score increased, the amount of false-positives (those identified as faking by the indicator score that did not change their scores substantially) was expected to decrease.

I then examined how this method (at these respective cut-scores) impacted hiring decisions in multiple select-in and select-out contexts. The same method was used to examine faking on the relevant predictors of Conscientiousness and Neuroticism scores respectively, as well as for individuals that were found to fake on both scales. First, I

created four groups of applicants based on the faking indicator scores for the various predictors (all applicants, applicants with indicator scores above a cut-score of 0 removed, applicants with indicator scores above a cut-score of 1 removed, and applicants with indicator scores above a cut-score of 2 removed).

To examine impact on select-in decisions, I then compared the all-applicants group with each of the groups that had applicants removed based on cut-scores respectively for the top 5%, 10%, 20%, and 30% of scorers. These percentages were chosen based on similar analyses reported in the extant literature (Mueller-Hanson et al., 2003; Peterson et al., 2009; Rosse et al., 1998). The improvements made (upon displacement of honest responders and the proportion of fakers hired) by using the method at various cut-scores, along with the rate of false positives, were examined for each of the aforementioned select-in rates.

False positive faking identification as a result of this method was examined by identifying the proportion of honest responders (as defined using the established confidence intervals) that would be removed from consideration due to faking indicator scores above the various cut-scores established. To examine the impact of this method of faking detection on select-out decisions, the number of honest respondents in the applicant condition that were below the threshold due to displacement from individuals identified as fakers (that the method identified as fakers at various cut-scores) that were above the threshold was counted. Thresholds for selection were compared at 70%, 50%, and 30%. These values were chosen to provide a range relevant for the majority of

applied contexts (aside from those involving extreme selectivity or extreme permissibility) as described in Berry and Sackett (2009).

Finally, for each independent context (the entire sample, select-in, select-out, curvilinear selection, and across all of these contexts combined) the respective indicators were compared using the raw values for correct faking identifications and false positive classifications, as well as with a single combined measure of the two (represented with correct decision proportions) for overall performance. Then, paired-samples t-tests were conducted to further compare the respective indicators, independently for all three of the aforementioned criteria and for each context. As multiple t-tests were conducted, exact p-values and effect sizes (for each independent analysis) are presented for researchers concerned with an increased possibility of Type I errors.

CHAPTER V

RESULTS

Descriptive Statistics

Reliabilities

Reliabilities for the sample's NEO-PI-R scores were calculated using Cronbach's alpha in the statistical program SPSS. In the research condition, the five factors evidenced Cronbach's alphas that ranged from .85 (Openness) to .91 (Neuroticism), with Conscientiousness ($\alpha = .90$), Neuroticism ($\alpha = .91$), and Extraversion ($\alpha = .88$) being the three highest. Cronbach's alphas for the individual facets under Conscientiousness for the research condition ranged from .58 (Achievement Striving) to .76 (Deliberation) with all facets other than Achievement Striving ($\alpha = .58$) evidencing Cronbach's alphas $> .67$. Cronbach's alphas for the individual facets under Neuroticism for the research condition were slightly higher, ranging from .70 (Impulsiveness) to .78 (Depression). Cronbach's alphas for the individual facets under Extraversion for the research condition were similar, ranging from .67 (Excitement-Seeking) to .78 (Assertiveness). These figures are consistent with previous research in both Romanian and non-Romanian samples.

In the applicant condition, the five factors evidenced Cronbach's alphas that ranged from .79 (Openness) to .89 (Neuroticism), with Conscientiousness ($\alpha = .88$), Neuroticism ($\alpha = .89$), and Extraversion ($\alpha = .85$), again being the three highest. Cronbach's alphas for the individual facets under Conscientiousness for the applicant

condition ranged from .70 (Order) to .81 (Achievement Striving). Cronbach's alphas for the individual facets under Neuroticism for the applicant condition ranged from .72 (Self-Consciousness) to .79 (Anxiety). Cronbach's alphas for the individual facets under Extraversion for the applicant condition ranged from .73 (Positive Emotions) to .78 (Warmth). Again, these figures are consistent with previous research. Test-retest reliabilities were .92 for Conscientiousness, .93 for Neuroticism, and .92 for Extraversion.

Correlations Between Research Factor Scores and Faking Indicators

In an attempt to ascertain whether the Kuncel & Borneman (2007) approach to faking detection remained (as reported in their original publication) uncorrelated with personality outside of the lab setting, I also analyzed the sample's correlations between the five respective factors' results from the research condition and the respective faking indicator scores (quantitative and qualitative). The quantitative faking indicator score was not significantly correlated with Neuroticism ($r[211] = -.01, p = .87$), Extraversion, ($r[211] = .02, p = .81$), Openness to Experience ($r[211] = .07, p = .35$), Agreeableness ($r[211] = -.08, p = .27$), nor with Conscientiousness, $r(211) = -.00, p = .95$. The qualitative faking indicator score, however, was highly significantly correlated with Neuroticism ($r[211] = .39, p \leq .0005$), Agreeableness ($r[211] = -.26, p < .0005$), and Conscientiousness, $r(211) = -.33, p \leq .0005$. Further, the qualitative faking indicator was significantly correlated with Extraversion, ($r[211] = -.16, p = .02$), however, it was not significantly correlated with Openness to Experience $r(211) = -.05, p = .44$. Table 5 presents these results.

Table 5. *Correlations Between NEO-PI-R Factor Results from the Research Condition and the Respective Faking Indicator Scores (Quantitative and Qualitative).*

Fact	<i>M</i>	<i>SD</i>	N	E	O	A	C	Qn	Ql
N	82.89	20.83	--						
E	110.67	18.40	-0.34**	--					
O	108.16	16.10	0.07	0.41**	--				
A	112.65	17.52	-0.36**	0.10	0.19**	--			
C	121.32	20.15	-0.57**	0.37**	0.13	0.45**	--		
Qn	12.05	8.67	-0.01	0.02	0.07	-0.08	0.00	--	
Ql	4.81	5.50	0.39**	-0.16*	-0.05	-0.26**	-0.33**	0.00	--

Note. *M* represents the mean of the sample's scores for the respective factors. *SD* represents the standard deviation of those scores. Split cells are divided such that Pearson's correlation coefficient (*r*) is presented above the line and the significance level (*p*) is presented below the line. ** denotes $p < .01$ and * denotes $p < .05$.

Factor Score Changes (Between Applicant and Research Conditions)

A series of paired-samples t-tests was also conducted to analyze score changes (between the applicant and research condition) for the respective personality factors. The

213 participants had an average factor-level Neuroticism score change of -3.35 ($SD = 7.89$), indicating a highly significant score decrease, $t(212) = -6.22, p \leq .0005, d = -0.43$. The 213 participants had an average factor-level Extraversion score change of 2.25 ($SD = 7.44$), indicating a highly significant score increase, $t(212) = 4.41, p \leq .0005, d = 0.30$. The 213 participants had an average factor-level Openness to Experience score change of -1.81 ($SD = 6.39$), indicating a highly significant score decrease, $t(212) = -4.14, p \leq .0005, d = -0.28$. The 213 participants had an average factor-level Agreeableness score change of 0.23 ($SD = 7.46$), indicating that there was no significant score change, $t(212) = 0.45, p = .65, d = 0.03$. The 213 participants had an average factor-level Conscientiousness score change of 6.41 ($SD = 7.95$), indicating a highly significant score increase, $t(212) = 11.78, p \leq .0005, d = 0.81$. Table 6 presents the means and standard deviations of scores for each factor from the respective conditions and for the difference scores (between conditions), as well as the 95% confidence interval (upper and lower boundary), t-statistic, significance level, and effect size for the paired-samples tests.

Table 6. *Paired-Samples t-Test Results for Differences Between Conditions for Each of the Five Personality Factors, Along with Means and Standard Deviations from the Respective Conditions.*

Factor	<i>M</i>	<i>SD</i>	<i>M_D</i>	<i>SD_D</i>	<i>LCI_D</i>	<i>UCI_D</i>	<i>t_D</i>	<i>p</i>	<i>d</i>
	82.89	20.83							
N	79.54	18.69	-3.35	7.87	-4.41	-2.29	-6.22	.00**	-0.43
	110.67	18.40							
E	112.92	15.97	2.25	7.44	1.24	3.25	4.41	.00**	0.30
	108.16	16.10							
O	106.35	15.38	-1.81	6.39	-2.68	-0.95	-4.14	.00**	-0.28
	112.65	17.52							
A	112.88	15.31	0.23	7.46	-0.78	1.24	0.45	.65	0.03
	121.32	20.15							
C	127.73	16.72	6.41	7.95	5.34	7.48	11.77	.00**	0.81

Note. Split cells are divided such that the research condition is presented above the line and the applicant condition is presented below the line. *M_D* represents the mean difference (from research to applicant) between conditions, *SD_D* represents the standard deviation of those differences, *LCI_D* and *UCI_D* represent the lower and upper boundaries of the 95% confidence interval for the mean differences respectively, *t_D* represents the t-statistic for the paired sample test of mean differences between conditions, *p* represent the significance level of those t-statistics, and *d* represents the effect size (with positive values representing an increase from the research context to the application context). ** denotes $p < .01$.

Research Question 2

To assess the utility of this method in a real-world application context, I examined the ability of the method to identify individuals categorized as true fakers (respectively for the 1 SD and ½ SD methods) at three cut-scores (0, 1, and 2 standard deviations above the mean faking indicator score) for each predictor.

1 SD Categorization Method

For Conscientiousness, my quantitative faking indicator correctly identified 54% (15/28) of fakers above the mean indicator score, while resulting in 87 false positive identifications, for an approximate correct decision proportion of $p = .53$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 22% (6/28) of fakers, while resulting in 29 false positives, for an approximate correct decision proportion of $p = .76$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 7% (2/28) of fakers, while resulting in four false positives, for an approximate correct decision proportion of $p = .86$.

For Neuroticism, the quantitative faking indicator correctly identified approximately 58% (19/33) of fakers above the mean indicator score, while resulting in 87 false positive identifications, for an approximate correct decision proportion of $p = .53$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 18% (6/33) of fakers, while resulting in 28 false positives, for an approximate correct decision proportion of $p = .74$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 12% (4/33) of fakers, while resulting in two false positives, for an approximate correct decision proportion of $p = .85$.

For Extraversion, the quantitative faking indicator correctly identified approximately 60% (32/53) of fakers above the mean indicator score, while resulting in 70 false positive identifications, for an approximate correct decision proportion of $p = .57$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 15% (8/53) of fakers, while resulting in 25 false positives, for an

approximate correct decision proportion of $p = .67$. At 2 SD above the mean, the quantitative indicator correctly identified less than 1% (3/53) of fakers, while resulting in three false positives, for an approximate correct decision proportion of $p = .75$. Table 7 presents these results.

Table 7. *1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Quantitative Faking Indicator.*

		Cut-Score		
Predictor	Results	>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	15/28	6/28	2/28
	False Positives	87	29	4
Neuroticism	Correct Faker Identifications	19/33	6/33	4/33
	False Positives	87	28	2
Extraversion	Correct Faker Identifications	32/53	8/53	3/53
	False Positives	70	25	3

Note. Fakers identified are listed as a ratio of those caught and those present. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

½ SD Categorization Method

For Conscientiousness, my quantitative faking indicator correctly identified approximately 54% (36/67) of fakers above the mean indicator score, while resulting in 67 false positive identifications, for an approximate correct decision proportion of $p = .54$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 19% (13/67) of fakers, while resulting in 20 false positives, for an

approximate correct decision proportion of $p = .65$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 4% (3/67) of fakers, while resulting in three false positives, for an approximate correct decision proportion of $p = .69$.

For Neuroticism, my quantitative faking indicator correctly identified approximately 55% (23/42) of fakers above the mean indicator score, while resulting in 89 false positive identifications, for an approximate correct decision proportion of $p = .49$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 21% (9/42) of fakers, while resulting in 28 false positives, for an approximate correct decision proportion of $p = .71$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 12% (5/42) of fakers, while resulting in just one false positive, for an approximate correct decision proportion of $p = .82$.

As mentioned previously, for Extraversion the respective categorization methods (1 SD and $\frac{1}{2}$ SD) resulted in the same decisions, therefore all results for Extraversion are identical and are not repeated in text. Readers may refer to the previous section for this elaboration. Table 8 presents these results.

Table 8. $\frac{1}{2}$ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Quantitative Faking Indicator.

Predictor	Results	Cut-Score		
		>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	36/67	13/67	3/67
	False Positives	67	20	3
Neuroticism	Correct Faker Identifications	23/42	9/42	5/42
	False Positives	89	28	1
Extraversion	Correct Faker Identifications	32/53	8/53	3/53
	False Positives	70	25	3

Note. Fakers identified are listed as a ratio of those caught and those present. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Research Question 3

To examine the impact of my changes to the Kuncel and Borneman (2007) approach, I attempted to re-create their qualitative approach to the scoring scheme, allowing for a comparison between the results from that and those of my own quantitative technique. I examined the ability of their method to identify those individuals categorized as true fakers (respectively for the 1 SD and $\frac{1}{2}$ SD categorization methods) at the same three cut-scores, (0, 1, and 2 standard deviations above the mean faking indicator score) for each predictor.

1 SD Categorization Method

For Conscientiousness, the Kuncel and Borneman (2007) qualitative faking indicator correctly identified approximately 46% (13/28) of fakers above the mean

indicator score, while resulting in 95 false positive identifications, for an approximate correct decision proportion of $p = .48$. At 1 SD above the mean, the qualitative indicator correctly identified approximately 18% (5/28) of fakers, while resulting in 25 false positives, for an approximate correct decision proportion of $p = .77$. At 2 SD above the mean, the qualitative indicator correctly identified approximately 11% (3/28) of fakers, while resulting in three false positives, for an approximate correct decision proportion of $p = .87$.

For Neuroticism, the qualitative faking indicator correctly identified approximately 58% (19/33) of fakers above the mean indicator score, while resulting in 89 false positive identifications, for an approximate correct decision proportion of $p = .52$. At 1 SD above the mean, the qualitative indicator correctly identified approximately 15% (5/33) of fakers, while resulting in 24 false positives, for an approximate correct decision proportion of $p = .76$. At 2 SD above the mean, the qualitative indicator correctly identified approximately 9% (3/33) of fakers, while resulting in three false positives, for an approximate correct decision proportion of $p = .85$.

For Extraversion, the qualitative faking indicator correctly identified approximately 58% (31/53) of fakers above the mean indicator score, while resulting in 77 false positive identifications, for an approximate correct decision proportion of $p = .54$. At 1 SD above the mean, the qualitative indicator correctly identified approximately 13% (7/53) of fakers, while resulting in 22 false positives, for an approximate correct decision proportion of $p = .68$. At 2 SD above the mean, the qualitative indicator correctly identified approximately 6% (3/53) of fakers, while resulting in three false

positives, for an approximate correct decision proportion of $p = .75$. Table 9 presents these results.

Table 9. *1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Kuncel and Borneman (2007) Qualitative Faking Indicator.*

		Cut-Score		
Predictor	Results	>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	13/28	5/28	3/28
	False Positives	95	25	3
Neuroticism	Correct Faker Identifications	19/33	5/33	3/33
	False Positives	89	24	3
Extraversion	Correct Faker Identifications	32/53	7/53	3/53
	False Positives	77	22	3

Note. Fakers identified are listed as a ratio of those caught and those present. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

½ SD Categorization Method

For Conscientiousness, the Kuncel and Borneman (2007) qualitative faking indicator correctly identified approximately 51% (34/67) of fakers above the mean indicator score, while resulting in 74 false positive identifications, for an approximate correct decision proportion of $p = .50$. At 1 SD above the mean, the quantitative indicator correctly identified approximately 15% (10/67) of fakers, while resulting in 18 false positives, for an approximate correct decision proportion of $p = .65$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 6% (4/67) of

fakers, while resulting in two false positives, for an approximate correct decision proportion of $p = .69$.

For Neuroticism, the qualitative faking indicator correctly identified approximately 55% (23/42) of fakers above the mean indicator score, while resulting in 85 false positive identifications, for an approximate correct decision proportion of $p = .51$. At 1 SD above the mean, the quantitative indicator correctly identified 17% (7/42) of fakers, while resulting in 22 false positives, for an approximate correct decision proportion of $p = .73$. At 2 SD above the mean, the quantitative indicator correctly identified approximately 10% (4/42) of fakers, while resulting in two false positives, for an approximate correct decision proportion of $p = .81$.

As before, for Extraversion the respective categorization methods (1 SD and $\frac{1}{2}$ SD) resulted in the same decisions, therefore all results for Extraversion are identical and are not repeated in text. Readers may refer to the previous section for this elaboration. Table 10 presents these results.

Table 10. $\frac{1}{2}$ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Using the Kuncel and Borneman (2007) Qualitative Faking Indicator.

Predictor	Results	Cut-Score		
		>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	34/67	10/67	4/67
	False Positives	74	18	2
Neuroticism	Correct Faker Identifications	23/42	7/42	4/42
	False Positives	85	22	2
Extraversion	Correct Faker Identifications	32/53	7/53	3/53
	False Positives	77	22	3

Note. Fakers identified are listed as a ratio of those caught and those present. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

To further facilitate direct comparisons of the respective faking identification methods (quantitative vs. qualitative), the actual differences between the number of fakers identified and the number of false positives for the respective methods are presented in Table 11 and Table 12.

Table 11. *Differences in 1 SD Categorized Faker Identifications and False Positives at Various Cut-Scores Between my Quantitative Faking Indicator and the Kuncel and Borneman (2007) Qualitative Indicator.*

Predictor	Results	Cut-Score		
		>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	+2	+1	-1
	False Positives	-8	+4	+1
Neuroticism	Correct Faker Identifications	0	+1	+1
	False Positives	-2	+4	-1
Extraversion	Correct Faker Identifications	+0	+1	0
	False Positives	-7	+3	0

Note. Differences are presented in terms of increase or decrease from the qualitative method to the quantitative method. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Table 12. *Differences in ½ SD Categorized Faker Identifications and False Positives at Various Cut-Scores Between my Quantitative Faking Indicator and the Kuncel and Borneman (2007) Qualitative Indicator.*

Predictor	Results	Cut-Score		
		>M	1SD>M	2SD>M
Conscientiousness	Correct Faker Identifications	+2	+3	-1
	False Positives	-7	+2	+1
Neuroticism	Correct Faker Identifications	0	+2	0
	False Positives	+4	+6	+1
Extraversion	Correct Faker Identifications	+0	+1	0
	False Positives	-7	+3	0

Note. Differences are presented in terms of increase or decrease from the qualitative method to the quantitative method. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Paired-samples t-tests were also conducted to examine the differences in correct faking identifications, false-positive faking identifications, and correct decision proportions between the respective faking indicator methods. For 18 comparisons and the entire sample, the difference in the number of correctly identified fakers between the quantitative method ($M = 12.61$, $SD = 11.16$) and the qualitative method ($M = 11.89$, $SD = 11.07$) was significant, $t(17) = 2.85$, $p = .011$, $d = 0.67$. However, there was no significant difference in the number of false-positive faking identifications between the quantitative method ($M = 35.61$, $SD = 33.10$) and the qualitative method ($M = 35.89$, $SD = 35.43$), $t(17) = -0.27$, $p = .79$, $d = -0.06$. Finally, there was no significant difference between correct decision proportions for the quantitative method ($M = 0.68$, $SD = 0.12$) and the qualitative method ($M = 0.67$, $SD = 0.13$), $t(17) = 0.95$, $p = .36$, $d = 0.22$.

Research Question 4

To examine the impact of this method of faking detection on select-in decisions, comparisons were made between the top scorers in the applicant condition after having removed those individuals identified as fakers (at various cut-scores) and the top scorers without removing such individuals. These comparisons were made at selection rates of 10%, 20%, and 30% (or the value closest to these percentages as was possible given the data). The rate of false positives at these percentages was also observed, as were contrasts between the respective scoring schemes and true faking categorization methods.

Conscientiousness/ 1 SD

Using the 1 SD method of true faking categorization for Conscientiousness, the quantitative faking indicator identified approximately 43% (3/7) of fakers scoring in the

top 30% ($N = 64$), while resulting in 14 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .72$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator identified approximately 29% (2/7) of fakers scoring in the top 30%, while resulting in 16 false positives, for an approximate correct decision proportion of $p = .67$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified zero fakers scoring in the top 30%, while resulting in two false positives, for an approximate correct decision proportion of $p = .86$. At 1 SD the qualitative indicator also identified zero fakers scoring in the top 30%, while resulting in three false positives, for an approximate correct decision proportion of $p = .84$. At a cut-score of 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 30%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .89$.

Continuing, the quantitative faking indicator identified approximately 67% (2/6) of fakers scoring in the top 20.2% ($N = 43$), while resulting in 10 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .67$. At the same cut-score, the qualitative indicator identified approximately 17% (1/6) of fakers scoring in the top 20.2%, while resulting in 12 false positives, for an approximate correct decision proportion of $p = .60$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified zero fakers scoring in the top 20.2%, while resulting in one false positive, for an approximate correct decision proportion of $p = .84$. At 1 SD the qualitative indicator also identified zero fakers scoring in the top 20.2%, while resulting in two false positives, for

an approximate correct decision proportion of $p = .81$. At a cut-score of 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 20.2%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .86$.

Finally, the quantitative faking indicator did not identify fakers (0/1) scoring in the top 10.3% ($N = 22$), while resulting in five false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .72$. At the same cut-score, the qualitative indicator also did not identify fakers (0/1) scoring in the top 10.3%, while also resulting in five false positives, for an approximate correct decision proportion of $p = .72$. At cut-scores of 1 and 2 SD above the mean faking indicator score, neither faking indicator identified fakers (0/1) scoring in the top 10.3%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .95$. Table 13 presents these results.

Table 13. *Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Conscientiousness.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	0/1 (5)	0/1 (0)	0/1 (0)
	20%	2/6 (10)	0/6 (1)	0/6 (0)
	30%	3/7 (14)	0/7 (2)	0/7 (0)
Qualitative	10%	0/1 (5)	0/1 (0)	0/1 (0)
	20%	1/6 (12)	0/6 (2)	0/6 (0)
	30%	2/7 (16)	0/7 (3)	0/7 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Conscientiousness/ ½ SD

Using the ½ SD method of true faking categorization for Conscientiousness, the quantitative faking indicator identified 43% (6/14) of fakers scoring in the top 30% ($N = 64$), while resulting in 12 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .69$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator identified 36% (5/14) of fakers scoring in the top 30%, while resulting in 13 false positives, for an approximate correct decision proportion of $p = .66$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified zero fakers scoring in the top 30%, while resulting in two false positives, for a correct decision proportion of $p = .75$. At 1 SD the qualitative indicator also identified zero fakers scoring in the top 30%, while resulting in three false positives, for an approximate correct decision

proportion of $p = .73$. At a cut-score of 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 30%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .78$.

Continuing, the quantitative faking indicator identified approximately 27% (3/11) of fakers scoring in the top 20.2% ($N = 43$), while resulting in 10 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .58$. At the same cut-score, the qualitative indicator identified approximately 18% (2/11) of fakers scoring in the top 20.2%, while resulting in 11 false positives, for an approximate correct decision proportion of $p = .53$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified zero fakers scoring in the top 20.2%, while resulting in one false positive, for an approximate correct decision proportion of $p = .72$. At 1 SD the qualitative indicator also identified zero fakers scoring in the top 20.2%, while resulting in two false positives, for an approximate correct decision proportion of $p = .70$. At a cut-score of 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 20.2%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .74$.

Finally, the quantitative faking indicator identified approximately 17% (1/6) of fakers scoring in the top 10.3% ($N = 22$), while resulting in five false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .55$. At the same cut-score, the qualitative indicator also identified approximately 17% (1/6) of fakers scoring in the top 10.3%, while

resulting in four false positives, for an approximate correct decision proportion of $p = .59$. At cut-scores of 1 and 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 10.3%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .73$. Table 14 presents these results.

Table 14. *Impact on Select-In Decisions, when Using 1/2 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Conscientiousness.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	1/6 (5)	0/6 (0)	0/6 (0)
	20%	3/11 (10)	0/11 (1)	0/11 (0)
	30%	6/14 (12)	0/14 (2)	0/14 (0)
Qualitative	10%	1/6 (4)	0/6 (0)	0/6 (0)
	20%	2/11 (12)	0/11 (2)	0/11 (0)
	30%	5/14 (13)	0/14 (3)	0/14 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Neuroticism/ 1 SD

Using the 1 SD method of true faking categorization for Neuroticism, the quantitative faking indicator identified approximately 43% (3/7) of fakers scoring in the top 30.5% ($N = 65$), while resulting in 16 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .69$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator

also identified approximately 43% (3/7) of fakers scoring in the top 30.5%, while resulting in 15 false positives, for an approximate correct decision proportion of $p = .71$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 14% (1/7) fakers scoring in the top 30.5%, while resulting in two false positives, for an approximate correct decision proportion of $p = .88$. At 1 SD the qualitative indicator also identified approximately 14% (1/7) fakers scoring in the top 30.5%, while resulting in three false positives, for an approximate correct decision proportion of $p = .86$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 14% (1/7) fakers scoring in the top 30.5%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .91$. At 2 SD the qualitative indicator identified zero fakers scoring in the top 30.5%, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .89$.

Continuing, the quantitative faking indicator identified approximately 33% (1/3) of fakers scoring in the top 20.2% ($N = 43$), while resulting in 11 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .70$. At the same cut-score, the qualitative indicator identified approximately 67% (2/3) of fakers scoring in the top 20.2%, while resulting in nine false positives, for an approximate correct decision proportion of $p = .77$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator also identified approximately 33% (1/3) of fakers scoring in the top 20.2%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .95$. At 1 SD

the qualitative indicator also identified approximately 33% (1/3) of fakers scoring in the top 20.2%, while resulting in two false positives, for an approximate correct decision proportion of $p = .91$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 33% (1/3) fakers scoring in the top 20.2%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .95$. At 2 SD the qualitative indicator identified zero fakers scoring in the top 20.2%, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .93$.

Finally, the quantitative faking indicator identified 50% (1/2) of fakers scoring in the top 9.4% ($N = 20$), while resulting in six false positives at a cut-score of anything above the sample's mean faking indicator score, for a correct decision proportion of $p = .65$. At the same cut-score, the qualitative indicator also identified 50% (1/2) of fakers scoring in the top 9.4%, while resulting in five false positives, for a correct decision proportion of $p = .70$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified 50% (1/2) of fakers scoring in the top 9.4%, while resulting in zero false positives, for a correct decision proportion of $p = .95$. At 1 SD the qualitative indicator also identified 50% (1/2) of fakers scoring in the top 9.4%, while resulting in one false positive, for a correct decision proportion of $p = .90$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 50% (1/2) of fakers scoring in the top 9.4%, while resulting in zero false positives, for a correct decision proportion of $p = .95$. At 2 SD the qualitative indicator identified zero fakers scoring in the top 9.4%, while also resulting in zero false positives, for a correct decision proportion of $p = .90$. Table 15 presents these results.

Table 15. *Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Neuroticism.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	1/2 (6)	1/2 (0)	1/2 (0)
	20%	1/3 (11)	1/3 (0)	1/3 (0)
	30%	3/7 (16)	1/7 (2)	1/7 (0)
Qualitative	10%	1/2 (5)	1/2 (1)	0/2 (0)
	20%	2/3 (9)	1/3 (2)	0/3 (0)
	30%	3/7 (15)	1/7 (3)	0/7 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Neuroticism/ ½ SD

Using the ½ SD method of true faking categorization for Neuroticism, the quantitative faking indicator identified approximately 33% (3/9) of fakers scoring in the top 30.5% ($N = 65$), while resulting in 20 false positives at a cut-score of anything above the sample's mean faking indicator score, for a correct decision proportion of $p = .60$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator also identified approximately 33% (3/9) of fakers scoring in the top 30.5%, while resulting in 15 false positives, for an approximate correct decision proportion of $p = .68$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 22% (2/9) of fakers scoring in the top 30.5%, while resulting in three false positives, for an approximate correct decision proportion of $p = .85$. At 1 SD the qualitative indicator identified approximately 11% (1/9) fakers scoring in the top 30.5%,

while also resulting in three false positives, for an approximate correct decision proportion of $p = .83$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 11% (1/9) fakers scoring in the top 30.5%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .88$. At 2 SD the qualitative indicator identified zero fakers scoring in the top 30.5%, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .86$.

Continuing, the quantitative faking indicator identified 50% (2/4) of fakers scoring in the top 20.2% ($N = 43$), while resulting in 14 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .63$. At the same cut-score, the qualitative indicator also identified 50% (2/4) of fakers scoring in the top 20.2%, while resulting in nine false positives, for an approximate correct decision proportion of $p = .74$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified 50% (2/4) of fakers scoring in the top 20.2%, while resulting in one false positive, for an approximate correct decision proportion of $p = .93$. At 1 SD the qualitative indicator identified 25% (1/4) of fakers scoring in the top 20.2%, while resulting in two false positives, for an approximate correct decision proportion of $p = .88$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified 25% (1/4) fakers scoring in the top 20.2%, while resulting in zero false

positives, for an approximate correct decision proportion of $p = .93$. At 2 SD the qualitative indicator identified zero fakers, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .91$.

Finally, the quantitative faking indicator identified approximately 67% (2/3) of fakers scoring in the top 9.4% ($N = 20$), while resulting in six false positives at a cut-score of anything above the sample's mean faking indicator score, for a correct decision proportion of $p = .65$. At the same cut-score, the qualitative indicator identified approximately 33% (1/3) of fakers scoring in the top 9.4%, while resulting in five false positives, for a correct decision proportion of $p = .65$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 67% (2/3) of fakers scoring in the top 9.4%, while resulting in one false positive, for a correct decision proportion of $p = .90$. At 1 SD the qualitative indicator identified approximately 33% (1/3) of fakers scoring in the top 9.4%, while also resulting in one false positive, for a correct decision proportion of $p = .85$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 33% (1/3) fakers scoring in the top 9.4%, while resulting in zero false positives, for a correct decision proportion of $p = .90$. At 2 SD the qualitative indicator identified zero fakers scoring in the top 9.4%, while also resulting in zero false positives, for a correct decision proportion of $p = .85$. Table 16 presents these results.

Table 16. *Impact on Select-In Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Neuroticism.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	2/3 (6)	2/3 (1)	1/3 (0)
	20%	2/4 (14)	2/4 (1)	1/4 (0)
	30%	3/9 (20)	2/9 (3)	1/9 (0)
Qualitative	10%	1/3 (5)	1/3 (1)	0/3 (0)
	20%	2/4 (9)	1/4 (2)	0/4 (0)
	30%	3/9 (15)	1/9 (3)	0/9 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Extraversion/ 1 SD

Using the 1 SD method of true faking categorization for Extraversion, the quantitative faking indicator identified approximately 44% (7/16) of fakers scoring in the top 30% ($N = 64$), while resulting in 15 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .63$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator identified approximately 50% (8/16) of fakers scoring in the top 30%, while resulting in 18 false positives, for an approximate correct decision proportion of $p = .59$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 13% (2/16) of fakers scoring in the top 30%, while resulting in four false positives, for an approximate correct decision proportion of $p = .72$. At 1 SD the qualitative indicator also identified approximately 13% (2/16) of fakers scoring in the top

30%, while resulting in five false positives, for an approximate correct decision proportion of $p = .70$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 6% (1/16) fakers scoring in the top 30%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .77$. At 2 SD the qualitative indicator identified approximately 13% (2/16) of fakers scoring in the top 30%, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .78$.

Continuing, the quantitative faking indicator identified 27% (3/11) of fakers scoring in the top 20.2% ($N = 43$), while resulting in nine false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .60$. At the same cut-score, the qualitative indicator identified 36% (4/11) of fakers scoring in the top 20.2%, while resulting in 13 false positives, for an approximate correct decision proportion of $p = .53$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified zero fakers scoring in the top 20.2%, while resulting in three false positives, for an approximate correct decision proportion of $p = .67$. At 1 SD the qualitative indicator also identified zero fakers scoring in the top 20.2%, while resulting in four false positives, for an approximate correct decision proportion of $p = .65$. At a cut-score of 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 20.2%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .74$.

Finally, the quantitative faking indicator identified 40% (2/5) of fakers scoring in the top 9.9% ($N = 21$), while resulting in two false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .76$. At the same cut-score, the qualitative indicator also identified 40% (2/5) of fakers scoring in the top 9.9%, while resulting in five false positives, for an approximate correct decision proportion of $p = .62$. At cut-scores of 1 and 2 SD above the mean faking indicator score, neither faking indicator identified fakers scoring in the top 9.9%, nor did they result in any false positives, leaving both with an approximate correct decision proportion of $p = .76$. Table 17 presents these results.

Table 17. *Impact on Select-In Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Extraversion.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	2/5 (2)	0/5 (0)	0/5 (0)
	20%	3/11 (9)	0/11 (3)	0/11 (0)
	30%	7/16 (15)	2/16 (4)	1/16 (0)
Qualitative	10%	2/5 (5)	0/5 (0)	0/5 (0)
	20%	4/11 (13)	0/11 (4)	0/11 (0)
	30%	8/16 (18)	2/16 (5)	2/16 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Extraversion/ ½ SD

As before, for Extraversion the respective categorization methods (1 SD and ½ SD) resulted in the same decisions, therefore all results for Extraversion are identical and are not repeated in text. Readers may refer to the previous section for this elaboration. Table 18 presents these results.

Table 18. Impact on Select-In Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Selection Rates for the Predictor Extraversion.

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	10%	2/5 (2)	0/5 (0)	0/5 (0)
	20%	3/11 (9)	0/11 (3)	0/11 (0)
	30%	7/16 (15)	2/16 (4)	1/16 (0)
Qualitative	10%	2/5 (5)	0/5 (0)	0/5 (0)
	20%	4/11 (13)	0/11 (4)	0/11 (0)
	30%	8/16 (18)	2/16 (5)	2/16 (0)

Note. Selection rates may be approximate. Fakers identified are listed as a ratio of those caught and those present. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Paired-samples t-tests were also conducted to examine the differences in correct faking identifications, false-positive faking identifications, and correct decision proportions between the respective faking indicator methods. For 54 comparisons made with select-in decisions, there was no significant difference in the number of correctly identified fakers between the quantitative method ($M = 1.33$, $SD = 1.66$) and the qualitative method ($M = 1.20$, $SD = 1.82$), $t(53) = 1.55$, $p = .13$, $d = 0.21$. However, the

difference in the number of false-positive faking identifications between the quantitative method ($M = 3.85$, $SD = 5.37$) and the qualitative method ($M = 4.40$, $SD = 5.48$) was marginally significant, $t(53) = -1.99$, $p = .052$, $d = -0.27$. Finally, the difference between correct decision proportions for the quantitative method ($M = 0.77$, $SD = 0.11$) and the qualitative method ($M = 0.75$, $SD = 0.11$) was highly significant, $t(53) = 2.67$, $p = .009$, $d = 0.36$.

Research Question 5

To examine the impact of this method of faking detection on select-out decisions, the number of honest respondents in the applicant condition that were below the threshold due to displacement as a result of faking was analyzed. This was done by counting the number of individuals above the threshold that were categorized (by the 1 SD and $\frac{1}{2}$ SD methods respectively) as true fakers and then contrasting that total number of displaced individuals with the number that were subsequently identified as fakers (by the respective indicators at the three cut-scores). This effectively offers insight toward the efficacy of this approach to mitigate the deleterious displacement effects of faking in select-out decisions. These contrasts were made at thresholds of 50% and 70% (or as close to these percentages as was reasonable given the data). The number of false positives above these thresholds was also recorded.

Conscientiousness/ 1 SD

Using the 1 SD method of true faking categorization for Conscientiousness, the quantitative faking indicator identified approximately 42% (5/12) of fakers scoring at or above a threshold of 50.7% ($N = 108$), while resulting in 36 false positives at a cut-score

of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .60$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator also identified approximately 33% (4/12) of fakers scoring at or above a threshold of 50.7%, while resulting in 39 false positives, for an approximate correct decision proportion of $p = .56$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 8% (1/12) of fakers scoring at or above a threshold of 50.7%, while resulting in nine false positives, for an approximate correct decision proportion of $p = .81$. At 1 SD the qualitative indicator also identified approximately 8% (1/12) of fakers scoring above a threshold of 50.7%, while also resulting in nine false positives, for an approximate correct decision proportion of $p = .81$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified zero fakers scoring at or above a threshold of 50.7%, while resulting in one false positive, for an approximate correct decision proportion of $p = .88$. At 2 SD the qualitative indicator identified approximately 8% (1/12) of fakers scoring at or above a threshold of 50.7%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .90$.

Continuing, the quantitative faking indicator identified approximately 44% (7/16) of fakers scoring at or above a threshold of 70.9% ($N = 151$), while resulting in 54 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .58$. At the same cut-score, the qualitative indicator identified approximately 38% (6/16) of fakers scoring at or above a threshold of 70.9%, while resulting in 61 false positives, for an approximate correct decision proportion of $p = .53$. At a cut-score of 1 SD above the mean faking indicator

score, the quantitative indicator identified approximately 13% (2/16) of fakers scoring at or above a threshold of 70.9%, while resulting in 17 false positives, for an approximate correct decision proportion of $p = .79$. At 1 SD the qualitative indicator also identified approximately 13% (2/16) of fakers scoring at or above a threshold of 70.9%, while resulting in 14 false positives, for an approximate correct decision proportion of $p = .81$.

At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 6% (1/16) of fakers scoring at or above a threshold of 70.9%, while resulting in two false positives, for an approximate correct decision proportion of $p = .89$. At 2 SD the qualitative indicator identified approximately 13% (2/16) of fakers scoring at or above a threshold of 70.9%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .91$. Table 19 presents these results.

Table 19. *Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Conscientiousness.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	5/12 (36)	1/12 (9)	0/12 (1)
	70%	7/16 (54)	2/16 (17)	1/16 (2)
Qualitative	50%	4/12 (39)	1/12 (9)	1/12 (0)
	70%	6/16 (61)	2/16 (14)	2/16 (0)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Conscientiousness/ ½ SD

Using the ½ SD method of true faking categorization for Conscientiousness, the quantitative faking indicator identified approximately 45% (13/29) of fakers scoring at or above a threshold of 50.7% ($N = 108$), while resulting in 29 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .58$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator also identified approximately 37% (11/29) of fakers scoring at or above a threshold of 50.7%, while resulting in 32 false positives, for an approximate correct decision proportion of $p = .54$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 14% (4/29) of fakers scoring at or above a threshold of 50.7%, while resulting in five false positives, for an approximate correct decision proportion of $p = .72$. At 1 SD the qualitative indicator also identified approximately 14% (4/29) of fakers scoring above a threshold of 50.7%, while resulting in six false positives, for an approximate correct decision proportion of $p = .71$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified zero fakers scoring at or above a threshold of 50.7%, while resulting in one false positive, for an approximate correct decision proportion of $p = .72$. At 2 SD the qualitative indicator identified approximately 3% (1/29) of fakers scoring at or above a threshold of 50.7%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .74$.

Finally, the quantitative faking indicator identified approximately 48% (22/46) of fakers scoring at or above a threshold of 70.9% ($N = 151$), while resulting in 40 false

positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .58$. At the same cut-score, the qualitative indicator also identified approximately 48% (22/46) of fakers scoring at or above a threshold of 70.9%, while resulting in 45 false positives, for an approximate correct decision proportion of $p = .54$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 15% (7/46) of fakers scoring at or above a threshold of 70.9%, while resulting in 10 false positives, for an approximate correct decision proportion of $p = .68$. At 1 SD the qualitative indicator identified approximately 13% (6/46) of fakers scoring at or above a threshold of 70.9%, while also resulting in 10 false positives, for an approximate correct decision proportion of $p = .67$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 2% (1/46) of fakers scoring at or above a threshold of 70.9%, while resulting in two false positives, for an approximate correct decision proportion of $p = .69$. At 2 SD the qualitative indicator also identified approximately 4% (2/46) of fakers scoring at or above a threshold of 70.9%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .71$. Table 20 presents these results.

Table 20. *Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Conscientiousness.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	13/29 (29)	4/29 (5)	0/29 (1)
	70%	22/46 (40)	7/46 (10)	1/46 (2)
Qualitative	50%	11/29 (32)	4/29 (6)	1/29 (0)
	70%	22/46 (45)	6/46 (10)	2/46 (0)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Neuroticism/ 1 SD

Using the 1 SD method of true faking categorization for Neuroticism, the quantitative faking indicator identified approximately 47% (7/15) of fakers scoring at or above a threshold of 50.7% ($N = 108$), while resulting in 29 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .66$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator also identified approximately 47% (7/15) of fakers scoring at or above a threshold of 50.7%, and also resulted in 29 false positives, for an approximate correct decision proportion of $p = .66$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 13% (2/15) of fakers scoring at or above a threshold of 50.7%, while resulting in six false positives, for an approximate correct decision proportion of $p = .82$. At 1 SD the qualitative indicator identified approximately 7% (1/15) of fakers scoring above a threshold of 50.7%, while

resulting in six false positives, for an approximate correct decision proportion of $p = .81$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 7% (1/15) fakers scoring at or above a threshold of 50.7%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .87$. At 2 SD the qualitative indicator identified zero fakers scoring at or above a threshold of 50.7%, while also resulting in zero false positives, for an approximate correct decision proportion of $p = .86$.

Continuing, the quantitative faking indicator identified approximately 58% (14/24) of fakers scoring at or above a threshold of 69% ($N = 147$), while resulting in 48 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .61$. At the same cut-score, the qualitative indicator identified approximately 54% (13/24) of fakers scoring at or above a threshold of 69%, while resulting in 47 false positives, for an approximate correct decision proportion of $p = .61$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 21% (5/24) of fakers scoring at or above a threshold of 69%, while resulting in 10 false positives, for an approximate correct decision proportion of $p = .80$. At 1 SD the qualitative indicator identified approximately 17% (4/24) of fakers scoring at or above a threshold of 69%, while resulting in nine false positives, for an approximate correct decision proportion of $p = .80$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 13% (3/24) of fakers scoring at or above a threshold of 69%, while resulting in zero false positives, for an approximate correct decision

proportion of $p = .86$. At 2 SD the qualitative indicator identified approximately 8% (2/24) of fakers scoring at or above a threshold of 69%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .85$. Table 21 presents these results.

Table 21. *Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Neuroticism.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	7/15 (29)	2/15 (6)	1/15 (0)
	70%	14/24 (48)	5/24 (10)	3/24 (0)
Qualitative	50%	7/15 (29)	1/15 (6)	0/15 (0)
	70%	13/24 (47)	4/24 (9)	2/24 (0)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Neuroticism/ ½ SD

Using the ½ SD method of true faking categorization for Neuroticism, the quantitative faking indicator identified approximately 42% (8/19) of fakers scoring at or above a threshold of 50.7% ($N = 108$), while resulting in 33 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .59$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator also identified approximately 42% (8/19) of fakers scoring at or

above a threshold of 50.7%, while resulting in 28 false positives, for an approximate correct decision proportion of $p = .64$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 16% (3/19) of fakers scoring at or above a threshold of 50.7%, while resulting in eight false positives, for an approximate correct decision proportion of $p = .78$. At 1 SD the qualitative indicator identified approximately 5% (1/19) of fakers scoring above a threshold of 50.7%, while resulting in six false positives, for an approximate correct decision proportion of $p = .78$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator also identified approximately 5% (1/19) fakers scoring at or above a threshold of 50.7%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .83$. At 2 SD the qualitative indicator identified zero fakers scoring at or above a threshold of 50.7%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .82$.

Finally, the quantitative faking indicator identified approximately 55% (17/31) of fakers scoring at or above a threshold of 69% ($N = 147$), while resulting in 50 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .56$. At the same cut-score, the qualitative indicator identified approximately 48% (15/31) of fakers scoring at or above a threshold of 69%, while resulting in 45 false positives, for an approximate correct decision proportion of $p = .59$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 23% (7/31) of fakers scoring at or above a threshold of 69%, while resulting in 11 false positives, for an approximate

correct decision proportion of $p = .76$. At 1 SD the qualitative indicator identified approximately 16% (5/31) of fakers scoring at or above a threshold of 69%, while resulting in eight false positives, for an approximate correct decision proportion of $p = .77$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 9% (3/31) of fakers scoring at or above a threshold of 69%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .81$. At 2 SD the qualitative indicator identified approximately 6% (2/31) of fakers scoring at or above a threshold of 69%, while resulting in zero false positives, for an approximate correct decision proportion of $p = .80$. Table 22 presents these results.

Table 22. *Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Neuroticism.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	8/19 (33)	3/19 (8)	1/19 (0)
	70%	17/31 (50)	7/31 (11)	3/31 (0)
Qualitative	50%	8/19 (28)	1/19 (6)	0/19 (0)
	70%	15/31 (45)	5/31 (8)	2/31 (0)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Extraversion/ 1 SD

Using the 1 SD method of true faking categorization for Extraversion, the quantitative faking indicator identified 50% (14/28) of fakers scoring at or above a threshold of 51.2% ($N = 109$), while resulting in 34 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .56$. At the same cut-score, the Kuncel and Borneman (2007) qualitative indicator identified approximately 54% (15/28) of fakers scoring at or above a threshold of 51.2%, while resulting in 37 false positives, for an approximate correct decision proportion of $p = .54$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 14% (4/28) of fakers scoring at or above a threshold of 51.2%, while resulting in 12 false positives, for an approximate correct decision proportion of $p = .67$. At 1 SD the qualitative indicator identified approximately 17% (5/28) of fakers scoring above a threshold of 51.2%, while resulting in 13 false positives, for an approximate correct decision proportion of $p = .67$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 3% (1/28) of fakers scoring at or above a threshold of 51.2%, while resulting in two false positives, for an approximate correct decision proportion of $p = .73$. At 2 SD the qualitative indicator identified approximately 7% (2/28) of fakers scoring at or above a threshold of 51.2%, while also resulting in two false positives, for an approximate correct decision proportion of $p = .74$.

Continuing, the quantitative faking indicator identified approximately 56% (20/36) of fakers scoring at or above a threshold of 71.4% ($N = 152$), while resulting in

49 false positives at a cut-score of anything above the sample's mean faking indicator score, for an approximate correct decision proportion of $p = .57$. At the same cut-score, the qualitative indicator identified approximately 58% (21/36) of fakers scoring at or above a threshold of 71.4%, while resulting in 53 false positives, for an approximate correct decision proportion of $p = .55$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 17% (6/36) of fakers scoring at or above a threshold of 71.4%, while also resulting in 19 false positives, for an approximate correct decision proportion of $p = .68$. At 1 SD the qualitative indicator identified approximately 19% (7/36) of fakers scoring at or above a threshold of 71.4%, while resulting in 19 false positives, for an approximate correct decision proportion of $p = .68$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 8% (3/36) of fakers scoring at or above a threshold of 71.4%, while resulting in three false positives, for an approximate correct decision proportion of $p = .76$. At 2 SD the qualitative indicator also identified approximately 8% (3/36) of fakers scoring at or above a threshold of 71.4%, while also resulting in three false positives, for an approximate correct decision proportion of $p = .76$. Table 23 presents these results.

Table 23. *Impact on Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Extraversion.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	14/28 (34)	4/28 (12)	1/28 (2)
	70%	20/36 (49)	6/36 (19)	3/36 (3)
Qualitative	50%	15/28 (37)	5/28 (13)	2/28 (2)
	70%	21/36 (53)	7/36 (19)	3/36 (3)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Extraversion/ ½ SD

As before, for Extraversion the respective categorization methods (1 SD and ½ SD) resulted in the same decisions, therefore all results for Extraversion are identical and are not repeated in text. Readers may refer to the previous section for this

elaboration. Table 24 presents these results.

Table 24. *Impact on Select-Out Decisions, when Using ½ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and Select-Out Thresholds for the Predictor Extraversion.*

Faking Indicator	Selection Rate	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	50%	14/28 (34)	4/28 (12)	1/28 (2)
	70%	20/36 (49)	6/36 (19)	3/36 (3)
Qualitative	50%	15/28 (37)	5/28 (13)	2/28 (2)
	70%	21/36 (53)	7/36 (19)	3/36 (3)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present above the respective thresholds. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Paired-samples t-tests were also conducted to examine the differences in correct faking identifications, false-positive faking identifications, and correct decision proportions between the respective faking indicator methods. For 36 comparisons made with select-out decisions, there was no significant difference in the number of correctly identified fakers between the quantitative method ($M = 6.39$, $SD = 6.21$) and the qualitative method ($M = 6.28$, $SD = 6.25$), $t(35) = 0.63$, $p = .54$, $d = 0.10$. There was also no significant difference in the number of false-positive faking identifications made between the quantitative method ($M = 17.75$, $SD = 17.74$) and the qualitative method ($M = 18.00$, $SD = 18.94$), $t(35) = -0.59$, $p = .56$, $d = -0.10$. Finally, there was no significant difference between correct decision proportions for the quantitative method ($M = 0.71$, $SD = 0.11$) and the qualitative method ($M = 0.70$, $SD = 0.12$), $t(35) = 0.86$, $p = .40$, $d = 0.14$.

Exploratory Curvilinear Analysis

Recent theory and research has increasingly suggested that there may be a curvilinear relation between personality factors and workplace criteria (Judge, Piccolo, & Kosalka, 2009; Kaiser & Hogan, 2011; Le, Oh, Robbins, Ilies, Holland, & Westrick, 2011). It may be that extreme levels of certain personality factors or traits, whether high or low, can have a detrimental impact on important work behaviors. Considering this possibility, as an exploratory analysis, I assessed the impact of this faking detection method (contrasting both faking indicators at the three cut-off scores) with both methods of true faking categorization (1 SD and $\frac{1}{2}$ SD) while selecting out the top 10% and the bottom 10% (or as close to these values as was possible given the data) for the respective predictors.

1 SD Categorization Method

For Conscientiousness, the quantitative faking indicator identified 56% (14/25) of fakers remaining in the sample ($N = 168$), after having removed the top 10.3% ($N = 22$) and the bottom 10.8% ($N = 23$), at a cut-score of anything above the sample's mean faking indicator score. The quantitative indicator resulted in 67 false positives at this cut-score, for an approximate correct decision proportion of $p = .54$. At the same cut-score, the qualitative indicator identified 48% (12/25) of fakers remaining in the sample, while resulting in 76 false positives, for an approximate correct decision proportion of $p = .47$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified 24% (6/25) of fakers remaining in the sample, while resulting in 22 false positives, for an approximate correct decision proportion of $p = .76$. At 1 SD the

qualitative indicator identified 16% (4/25) of fakers remaining in the sample, while resulting in 18 false positives, for an approximate correct decision proportion of $p = .77$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified 8% (2/25) of fakers remaining in the sample, while resulting in three false positives, for an approximate correct decision proportion of $p = .85$. At 2 SD the qualitative indicator identified 12% (3/25) of fakers remaining in the sample, while resulting in one false positive, for an approximate correct decision proportion of $p = .86$.

For Neuroticism, the quantitative faking indicator identified approximately 59% (16/27) of fakers remaining in the sample ($N = 171$), after having removed the top 9.4% ($N = 20$) and the bottom 10.3% ($N = 22$), at a cut-score of anything above the sample's mean faking indicator score. The quantitative indicator resulted in 71 false positives at this cut-score, for an approximate correct decision proportion of $p = .52$. At the same cut-score, the qualitative indicator identified approximately 56% (15/27) of fakers remaining in the sample, while resulting in 70 false positives, for an approximate correct decision proportion of $p = .52$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 19% (5/27) of fakers remaining in the sample, while resulting in 24 false positives, for an approximate correct decision proportion of $p = .73$. At 1 SD the qualitative indicator identified approximately 15% (4/27) of fakers remaining in the sample, while resulting in 19 false positives, for an approximate correct decision proportion of $p = .75$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 11% (3/27) of fakers remaining in the sample, while resulting in one false positive, for an

approximate correct decision proportion of $p = .85$. At 2 SD the qualitative indicator also identified approximately 11% (3/27) of fakers remaining in the sample, while resulting in two false positives, for an approximate correct decision proportion of $p = .85$.

For Extraversion, the quantitative faking indicator identified approximately 62% (23/37) of fakers remaining in the sample ($N = 170$), after having removed the top 9.9% ($N = 21$) and the bottom 10.3% ($N = 22$), at a cut-score of anything above the sample's mean faking indicator score. The quantitative indicator resulted in 62 false positives at this cut-score, for an approximate correct decision proportion of $p = .55$. At the same cut-score, the qualitative indicator also identified approximately 62% (23/37) of fakers remaining in the sample, while resulting in 67 false positives, for an approximate correct decision proportion of $p = .52$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 19% (7/37) of fakers remaining in the sample, while resulting in 23 false positives, for an approximate correct decision proportion of $p = .69$. At 1 SD the qualitative indicator also identified approximately 19% (7/37) of fakers remaining in the sample, while resulting in 21 false positives, for a correct decision proportion of $p = .70$. At a cut-score of 2 SD above the mean faking indicator score, both faking indicators identified approximately 8% (3/37) of fakers remaining in the sample, while resulting in three false positives, leaving both with an approximate correct decision proportion of $p = .78$. Table 25 presents these results.

Table 25. *Impact on Curvilinear Select-Out Decisions, when Using 1 SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and All Three Predictors.*

Faking Indicator	Predictor	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	Conscientiousness	14/25 (67)	6/25 (22)	2/25 (3)
	Neuroticism	16/27 (71)	5/27 (24)	3/27 (1)
	Extraversion	23/37 (62)	7/37 (23)	3/37 (3)
Qualitative	Conscientiousness	12/25 (76)	4/25 (18)	3/25 (1)
	Neuroticism	15/27 (70)	4/27 (19)	3/27 (2)
	Extraversion	23/37 (67)	7/37 (21)	3/37 (3)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present in the remaining sample. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

½ SD Categorization Method

For Conscientiousness, the quantitative faking indicator identified approximately 55% (31/56) of fakers remaining in the sample ($N = 168$), after having removed the top 10.3% ($N = 22$) and the bottom 10.8% ($N = 23$), at a cut-score of anything above the sample's mean faking indicator score. The quantitative indicator resulted in 50 false positives at this cut-score, for an approximate correct decision proportion of $p = .55$. At the same cut-score, the qualitative indicator also identified approximately 55% (31/56) of fakers remaining in the sample, while resulting in 57 false positives, for an approximate correct decision proportion of $p = .52$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 23% (13/56) of fakers remaining in the sample, while resulting in 13 false positives, for an approximate correct decision proportion of $p = .67$. At 1 SD the qualitative indicator identified approximately

18% (10/56) of fakers remaining in the sample, while resulting in 12 false positives, for an approximate correct decision proportion of $p = .65$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 5% (3/56) of fakers remaining in the sample, while resulting in two false positives, for an approximate correct decision proportion of $p = .67$. At 2 SD the qualitative indicator identified approximately 7% (4/56) of fakers remaining in the sample, while resulting in zero false positives, for an approximate correct decision proportion of $p = .69$.

For Neuroticism, the quantitative faking indicator identified approximately 56% (19/34) of fakers remaining in the sample ($N = 171$), after having removed the top 9.4% ($N = 20$) and the bottom 10.3% ($N = 22$), at a cut-score of anything above the sample's mean faking indicator score. The quantitative indicator resulted in 73 false positives at this cut-score, for an approximate correct decision proportion of $p = .49$. At the same cut-score, the qualitative indicator identified approximately 53% (18/34) of fakers remaining in the sample, while resulting in 67 false positives, for an approximate correct decision proportion of $p = .51$. At a cut-score of 1 SD above the mean faking indicator score, the quantitative indicator identified approximately 21% (7/34) of fakers remaining in the sample, while resulting in 23 false positives, for an approximate correct decision proportion of $p = .71$. At 1 SD the qualitative indicator identified approximately 17% (6/34) of fakers remaining in the sample, while resulting in 17 false positives, for an approximate correct decision proportion of $p = .74$. At a cut-score of 2 SD above the mean faking indicator score, the quantitative faking indicator identified approximately 12% (4/34) of fakers remaining in the sample, while resulting in zero false positives, for

an approximate correct decision proportion of $p = .82$. At 2 SD the qualitative indicator also identified approximately 12% (4/34) of fakers remaining in the sample, while resulting in one false positive, for an approximate correct decision proportion of $p = .82$.

As before, for Extraversion the respective categorization methods (1 SD and $\frac{1}{2}$ SD) resulted in the same decisions, therefore all results for Extraversion are identical and are not repeated in text. Readers may refer to the previous section for this elaboration.

Table 26 presents these results.

Table 26. *Impact on Curvilinear Select-Out Decisions, when Using $\frac{1}{2}$ SD Faker-Categorizations, for the Respective Faking Indicators at Various Cut-Scores and All Three Predictors.*

Faking Indicator	Predictor	Cut-Score		
		>M	1SD>M	2SD>M
Quantitative	Conscientiousness	31/56 (50)	13/56 (13)	3/56 (2)
	Neuroticism	19/34 (73)	7/34 (23)	4/34 (0)
	Extraversion	23/37 (62)	7/37 (23)	3/37 (3)
Qualitative	Conscientiousness	31/56 (57)	10/56 (12)	4/56 (0)
	Neuroticism	18/34 (67)	6/34 (17)	4/34 (1)
	Extraversion	23/37 (67)	7/37 (21)	3/37 (3)

Note. Select-out thresholds may be approximate. The effect of the method on displacement is represented as a ratio of fakers identified and fakers present in the remaining sample. False positives are listed in parentheses. >M represents individuals above the mean cut-score; 1SD>M represents individuals more than one standard deviation above the mean cut-score; 2SD>M represents individuals more than two standard deviations above the mean cut-score.

Paired-samples t-tests were also conducted to examine the differences in correct faking identifications, false-positive faking identifications, and correct decision proportions between the respective faking indicator methods. For 18 comparisons made

in curvilinear contexts, the difference in the number of correctly identified fakers between the quantitative method ($M = 10.50$, $SD = 8.68$) and the qualitative method ($M = 10.00$, $SD = 8.56$) was marginally significant, $t(17) = 2.03$, $p = .06$, $d = 0.48$. However, there was no significant difference in the number of false-positive faking identifications made between the quantitative method ($M = 29.17$, $SD = 27.21$) and the qualitative method ($M = 29.00$, $SD = 28.98$), $t(17) = 0.17$, $p = .87$, $d = 0.04$. Finally, there was no significant difference between correct decision proportions for the quantitative method ($M = 0.68$, $SD = 0.12$) and the qualitative method ($M = 0.68$, $SD = 0.13$), $t(17) = 0.34$, $p = .74$, $d = 0.08$.

Concluding these analyses, paired-samples t-tests were also conducted to examine the differences in correct faking identifications, false-positive faking identifications, and correct decision proportions between the respective faking indicator methods for all 126 comparisons (made with the entire sample, select-in decisions, select-out decisions, and the curvilinear selection system). The difference in the number of correctly identified fakers between the quantitative method ($M = 5.70$, $SD = 7.60$) and the qualitative method ($M = 5.44$, $SD = 7.48$) was highly significant, $t(125) = 3.22$, $p = .002$, $d = 0.29$. However, there was no significant difference in the number of false-positive faking identifications made between the quantitative method ($M = 15.98$, $SD = 22.25$) and the qualitative method ($M = 16.25$, $SD = 23.23$), $t(125) = -1.10$, $p = .28$, $d = -0.10$. Finally, the difference between correct decision proportions for the quantitative method ($M = 0.72$, $SD = 0.12$) and the qualitative method ($M = 0.72$, $SD = 0.12$) was highly significant, $t(125) = 2.89$, $p = .005$, $d = 0.26$.

CHAPTER VI

DISCUSSION

Summary of Findings

To begin a summary of the findings of the current study, a note of caution regarding their interpretation must be made salient. As was evidenced by the range of values found with the various true-faking categorization methods herein investigated, there is no certain method for determining whether an individual is actually faking. The various methods used to assess applicants' faking may result in differential outcomes. Such results support the findings of previous research, which has evidenced that the method of categorization that is chosen can considerably impact the conclusions reached through such analyses (Peterson, Griffith, Converse, & Gammon, 2011).

With this taken into consideration, the results from the current study's analyses reflect only the use of the 1 SD and ½ SD methods of faking categorization. Using these two methods for categorizing individuals likely to have faked on a personality inventory, the number of individuals in the sample that were categorized as fakers varied from around 13% to nearly one-third of the sample, depending upon the specific combination of predictor and categorization method. Additionally, these results indicated that more individuals (or a similar number in one case) faked on measures of Conscientiousness

and Extraversion than on Neuroticism. This lends some support to previous findings that individuals are able to fake for job-related traits (Kroger & Turnbull, 1975; Raymark & Tafero, 2009). This also suggests that applicants may have an implicit understanding of the importance of Conscientiousness-related traits (that may not be explicitly job-related) as much as hiring professionals do, and that they attempt to respond to such items in an appropriate manner.

Moreover, fakers were found to be among the top percentages of scorers for all three predictors (resulting in the displacement of honest responders), when using either included method of faking categorization. For Conscientiousness, the percentage of fakers out of those individuals scoring above the three cut-rates ranged from 5% to 14% when categorized with the 1 SD approach, and from 22% to 27% when categorized using the $\frac{1}{2}$ SD approach. For Neuroticism, the percentage of fakers out of those individuals scoring above the three cut-rates ranged from 7% to 20% when categorized with the 1 SD approach, and from 9% to 15% when categorized using the $\frac{1}{2}$ SD approach. For Extraversion, the percentage of fakers out of those individuals scoring above the three cut-rates ranged from 24% to 26% when categorized with either the 1 SD or the $\frac{1}{2}$ SD approach.

For Conscientiousness, the percentage of fakers out of those individuals scoring above the two select-out thresholds was 11% when categorized with the 1 SD approach, and ranged from 27% to 30% when categorized using the $\frac{1}{2}$ SD approach. For Neuroticism, the percentage of fakers out of those individuals scoring above the two select-out thresholds ranged from 14% to 16% when categorized with the 1 SD approach,

and from 18% to 21% when categorized using the $\frac{1}{2}$ SD approach. For Extraversion, the percentage of fakers out of those individuals scoring above the two select-out thresholds ranged from 24% to 26% when categorized with either the 1 SD or the $\frac{1}{2}$ SD approach.

Regarding the Kuncel and Borneman (2007) proposed method of faking detection, several important findings emerged from the current study. First, the method translated well to contexts outside of the exact situation in which the method was developed. More specifically, when limited to a measure that relies on only five response options, the necessary criteria for selecting items as useful for faking identification still emerged at a functional quantity. Also, even when constrained to include one specific job family, there was enough variance in responses to evidence the requisite disagreement between applicants as to the most desirable responses.

Finally, examining the efficacy of the method with real-world applicants (rather than students directed to fake in a lab-setting) resulted in the successful identification of notable percentages of fakers. Although there was some attenuation (from the original study, which reported correct faking identifications ranging from 62% to 78%) of the percentage of fakers correctly identified from the entire sample (ranging from 51% to 60% in the current study when viewing both types of faking indicator at the lowest cut-score of anything above the mean), the decline was not as steep as one might have expected when considering the transition to the current method of inquiry. For instance, individuals presumably faked to varying degrees (or not at all), as they were not explicitly instructed how (or whether or not) to do so. Additionally, unlike in the current study where true fakers had to be categorized using a method of estimation, in the

original study it was known who was faking and who was not. Both of these differences may serve to explain part of the decrement in percentages evidenced here.

Moreover, the indicator score identifications were applied only after the respective indicator scores for the sample were standardized. This was done partly to account for the notion (discussed later) that contextualization effects may have accounted for some changing of scores, but most likely not for the most egregious offenders. This is also believed to have resulted in faking identifications of more extreme fakers, and therefore represents a test of a conservative application of this technique. As a result, individuals were identified as faking only when they exceeded (to varying degrees, when considering the use of three cut-scores) the mean faking indicator score (quantitative $M = 12.05$, qualitative $M = 4.81$), whereas anything on the positive side of the unstandardized indicator was considered faking in the original publication. This process, even when considering only the lowest cut-score (as in the preceding paragraph), may also serve to explain some of the attenuation (from the original study) of the percentages of fakers correctly identified in the current study.

Extrapolating, the current study further expands the understanding of this method's utility by this very process. By examining its efficacy at multiple cut-scores, rather than simply above or below a neutral faking indicator score, the interaction between the percentage of identified fakers and the risk of false positives becomes clearer. As would be expected, as cut-scores became more conservative (1 or 2 SD $> M$), the method correctly identified consistently lower numbers of fakers. However, another expected (yet beneficial) effect was that the number of false-positive faking

identifications evidenced an inverse relationship with the cut-score as well. Both of these effects were relatively stable across all combinations of faking categorization methods and faking indicator scores.

Regarding the impact of the changes made to this method of faking detection, direct comparisons between the qualitative and quantitative approaches for the entire sample revealed small differences that were consistent and significant for faking identification, but inconsistent for avoiding false positive decisions. For the entire sample and out of 18 possible comparisons (three predictors by three cut-scores by two true faking categorization methods), the quantitative indicator resulted in a greater number (ranging from one to three more) of correct faking identifications in approximately 56% (10/18) of the comparisons, the same number in approximately 33% (6/18) of comparisons, and a smaller number (one less) in only approximately 11% (2/18) of comparisons. The quantitative indicator resulted in a smaller number of false-positives (from one to eight less) in approximately 39% (7/18) of the comparisons, the same number in approximately 11% (2/18) of comparisons, and a greater number (from one to six more) in approximately 50% (9/18) of comparisons. In summary, for the overall sample the quantitative indicator consistently correctly identified the same or a greater number of fakers, while numbers of false-positive decisions made were comparable.

The overall performance of the respective faking indicators (rather than analyzing correct faking identification and false positive identifications separately) can be similarly compared when viewing these percentages in terms of greater, equivalent, or lower correct decision proportions. For the entire sample and out of 18 possible comparisons,

the quantitative indicator resulted in a greater correct decision proportion in approximately 44% (8/18) of comparisons, the same proportion in approximately 11% (2/18) of comparisons, and a lower proportion in approximately 44% (8/18) of comparisons. In summary, for the entire sample the overall performance (judged by the proportion of correct decisions made) of the respective indicators was comparable.

Further extending the research regarding this method's utility, comparisons of both indicators (at multiple cut-scores) at various select-in percentages revealed small differences as well, but were more consistent for both relevant criteria. In 54 possible comparisons (three select-in percentages by three predictors by three cut-scores by two true faking categorization methods), the quantitative indicator correctly identified a greater number (one more) of fakers in approximately 26% (14/54) of the comparisons, the same number in approximately 61% (33/54) of comparisons, and a smaller number (one less) in approximately 13% (7/54) of comparisons. The quantitative indicator also resulted in a smaller number (from one to four less) of false-positives in approximately 41% (22/54) of the comparisons, the same number in approximately 46% (25/54) of comparisons, and a greater number (from one to five more) in only approximately 13% (7/54) of comparisons.

Although statistical analyses did not reveal a significant difference between the two methods for faking identification (this may have been due to the relatively small differences evidenced), the difference nearly reached marginal significance and did evidence a relatively healthy effect size. Therefore, viewing the number of comparisons in which the quantitative indicator was superior may lead to clearer conclusions in this

instance. In summary, these results indicated that at more stringent selection rates, the quantitative faking indicator more often correctly identified the same or an even greater number of fakers, while also consistently resulting in fewer numbers of false-positive decisions.

Comparing the overall performance of the respective faking indicators for select-in decisions resulted in even more convincing findings. For select-in decisions and out of 54 possible comparisons, the quantitative indicator resulted in a greater correct decision proportion in approximately 56% (30/54) of comparisons, the same proportion in approximately 30% (16/54) of comparisons, and a lower proportion in approximately 15% (8/54) of comparisons. In summary, for the more stringent select-in decisions the overall performance (judged by the proportion of correct decisions made) of the quantitative indicator was significantly and consistently superior.

Also extending the research into this method's utility, comparisons of both indicators (at multiple cut-scores) at various select-out thresholds again revealed small but inconsistent differences. In 36 possible comparisons (two select-out thresholds by three predictors by three cut-scores by two true faking categorization methods), the quantitative indicator correctly identified a greater number of fakers (one or two more) in approximately 39% (14/36) of the comparisons, the same number in 22% (8/36) of comparisons, and a smaller number (one less) in approximately 39% (14/36) of comparisons. The quantitative indicator also resulted in a smaller number (from one to seven less) of false-positives in approximately 31% (11/36) of the comparisons, the same number in approximately 39% (14/36) of comparisons, and a greater number (from one to

five more) in approximately 31% (11/36) of comparisons. In summary, these results indicated that at more lenient select-out thresholds the two methods were comparable in faking identification and in avoiding false-positive decisions.

Comparing the overall performance of the respective faking indicators for select-out decisions reveals similar results. For select-out decisions and out of 36 possible comparisons, the quantitative indicator resulted in a greater correct decision proportion in approximately 42% (15/36) of comparisons, the same proportion in 25% (9/36) of comparisons, and a lower proportion in approximately 33% (12/36) of comparisons. In summary, for select-out decisions the overall performance (judged by the proportion of correct decisions made) of the respective faking indicators was comparable.

In a final extension of the research regarding the utility of this method, exploring its functionality with a curvilinear selection system evidenced small differences that were somewhat consistent for faking identification, but inconsistent for avoiding false positive decisions. In 18 possible comparisons (three predictors by three cut-scores by two true faking categorization methods), the quantitative indicator correctly identified a greater number of fakers (from one to three more) in approximately 39% (7/18) of the comparisons, the same number in 50% (9/18) of comparisons, and a smaller number (one less) in only approximately 11% (2/18) of comparisons. The quantitative indicator also resulted in a smaller number of false-positives (from one to nine less) in approximately 33% (6/18) of the comparisons, the same number in approximately 11% (2/18) of comparisons, and a greater number (from one to six more) in approximately 56% (10/18) of comparisons. These results indicated that with a curvilinear selection system, the

quantitative indicator consistently made a greater number of correct faking identifications, although the two indicators performed comparably in avoiding false-positive decisions.

Comparing the overall performance of the respective faking indicators for the curvilinear system decisions evidences inconsistent results. For select-out decisions and out of 36 possible comparisons, the quantitative indicator resulted in a greater correct decision proportion in approximately 39% (7/18) of comparisons, the same proportion in 17% (3/18) of comparisons, and a lower proportion in approximately 44% (8/18) of comparisons. In summary, for the curvilinear system the overall performance (judged by the proportion of correct decisions made) of the respective faking indicators was comparable.

Viewed collectively, the quantitative indicator performed better than the qualitative indicator in approximately 36% (45/126) of the respective contexts analyzed regarding correct faking identifications, as well in approximately 44% (56/126), and not as well in 20% (25/126). Furthermore, the quantitative indicator performed better than the qualitative indicator in approximately 29% (37/126) of the respective contexts analyzed regarding false-positive decisions, as well in approximately 34% (43/126), and not as well in approximately 37% (46/126). Considering overall performance using correct decision proportions, the quantitative indicator performed better than the qualitative indicator in approximately 44% (55/126) of the respective contexts analyzed, as well in approximately 24% (30/126), and not as well in approximately 33% (41/126).

Figures 5 through 52 (in Appendix B) depict all of the comparisons mentioned in the preceding paragraphs. When considering these comparisons in their entirety, the quantitative indicator evidenced a significant advantage regarding faking identifications and overall performance (as evaluated using correct decision proportions), while there was no significant difference between the respective methods regarding the avoidance of false-positive decisions. It is also important to note that the quantitative indicator performed better for both respective criteria in select-in contexts, which are typical of most selection systems. Considering such findings, these results suggest that adopting a more refined recoding scheme that is based on quantitative analysis (as compared to using judgment alone) of item response distributions may produce preferable results regarding overall performance and the two most important criteria in faking detection research in typical selection contexts.

Strengths

This study is (to my knowledge) unique in faking research in that it assesses the displacement effects of faking at the individual level, using a within-subjects design and real-world job applicants. Additionally, this study analyzed several methods of true-faking categorization, highlighting the lack of a reliable approach for identifying this phenomenon. Further, the promising results of the Kuncel and Borneman (2007) approach to faking detection were investigated thoroughly, within myriad contexts, serving to elucidate the strengths and limitations of the approach. The current study, therefore, addressed several limitations of the original publication regarding this

innovative method of faking detection as well as those of previous studies that attempted to assess faking in more general terms.

For instance, similar research that previously attempted to assess the extent of faking has suffered from notable limitations. For example, Hogan et al.'s (2007) within-subjects design regarding faking relied on two applicant conditions, rather than an applicant (faking) condition and research (honest) condition. Although the authors' assumption was that the initial assessment did not include faking, it is quite possible that both assessments were influenced by intentional distortion. The authors did attempt to address that limitation, but they did so by resorting to a between-subjects design with their inclusion of a research condition.

Further, Ellingson et al.'s (2007) within-subjects design regarding applicant faking relied on a personality measure (California Psychological Inventory, or CPI) that utilizes a true/false response set that restricts the type of faking that may occur to diametrically opposed answers only. Applicants might be much less likely to completely reverse an answer than to simply shift it from one side of a neutral endorsement to the other, or to a slightly less extreme endorsement. Additionally, while the authors did account for the possibility of the passage of time affecting score changes with a design counterbalanced for order effects, they analyzed rank-order changes through correlation rather than at the individual level. While the correlation results may have suggested that faking did not significantly impact score changes beyond the effects of time, deleterious displacement at the individual level may still have occurred due to faking.

Limitations

Limitations in faking research may be necessarily manifold. As previously stated, while the generalizability and ecological validity of faking research is enhanced with the use of real-world applicants, there is no certain method for determining the individuals whom are actually faking in such contexts. Various methods for assessing applicant faking have met with differential outcomes, as evidenced by the varying numbers of faking categorizations made by the respective methods used in the current study. Such results support the findings of previous research, which has evidenced that the method of categorization that is chosen can considerably impact the conclusions reached through such analyses (Peterson, Griffith, Converse, & Gammon, 2011).

Another limitation of the current study is the use of a judgmental approach in selecting the items recoded to construct the faking indicator scores. While the limitation of the use of judgment in assigning the recoded values was addressed, due to the nature of this method the selection of items for recoding may necessarily require the use of judgment. When assessing the changing of responses and disagreement between conditions over multiple response options, a complex interaction of movement between response options occurs, such that simple analyses of skewness and kurtosis will not reveal the items that best demonstrate the necessary criteria. Therefore, as a post hoc, exploratory measure, a panel of raters was tasked with rating the degree to which each item represented a good or poor faking indicator.

The inter-rater reliabilities for the respective items offer a method with which to quantify this necessarily qualitative process. Not only can agreement as to the utility of

the item be established, but with advanced rater-training and a properly granular rating system, the ratings may be useful in rank-ordering the selected items as to their expected effectiveness as a faking indicator. For instance, those items with the highest inter-rater reliabilities that also corresponded at the highest rating level of a good indicator (seven, in this case), could be weighted more heavily than items that had lower inter-rater reliabilities that were still determined to be useful as faking indicators, or than items with high inter-rater reliabilities at lower (yet still useful) ratings (five).

The extent of time that lapsed between the research and applicant condition may also be of some concern to researchers. It could be argued that changes in scores that occurred between conditions may have been due to actual changes in the individuals' personality over time, rather than deliberate faking. Without controlling for such effects by implementing a counterbalanced approach to the respective assessment conditions, this possibility cannot be ignored. However, again I believe that the nature of the method of faking categorizations used (that serves to identify the most extreme changes in scores) should offer a buffer against this concern. Additionally, it seems unlikely that an individual's natural evolution of personality would result in changes that were always consistent with those items that evidence the sample's disagreement over the direction of the change (which are selected for use as indicators of faking). While an individual's score changes may indeed be the result of an evolution of their personality over time, for an individual to have been identified as a faker using this method of detection, they would have changed in a direction consistent with theoretical faking across 42 items. Although this certainly could have occurred, it seems largely implausible.

Further limitations include the use of a relatively small, Romanian sample of Communications majors that may not generalize to other cultures or job families. Additionally, the lack of alternative measures of individual differences such as cognitive ability and social desirability (included in the original study) prohibited the examination of the effect of such differences on an individual's ability to avoid detection (Kuncel & Borneman, 2007).

Implications for Practice

Practical implications of the current research are numerous. First, the Kuncel and Borneman (2007) method of faking detection may represent a viable alternative for flagging applicants suspected of engaging in faking behaviors. The results suggesting that the method remained functional when applied to a context that relied on a more common personality measure, real-world applicants, and a specific job family offer support for the further use, investigation, and refinement of the approach. Additionally, this method may be amenable to hiring decisions made in any field and while incorporating any of myriad measures of personality in the selection system.

Continuing, applicants were found to disagree on all five factors when identifying unusual items, including the aforementioned Conscientiousness and on the Extraversion factor specifically included as a predictor for its job relevance. The fact that this occurred for all factors, with a relatively straightforward measure using statement presentation, suggests that disagreement is not due simply to confusion or misunderstanding as to the meaning of an item. While Openness to Experience items were overrepresented in the subset of items selected as faking indicators, it does not

appear necessary to rely on these traditionally more ambiguous items alone when applying this method. It may even be that the accuracy of predictions increments as more ambiguous items such as Openness are avoided in favor of more straightforward or ostensibly job-related items.

Here, it is important to note that the items selected serve as indicators of faking behavior only, and are not used as a measure of faking respectively for the factors that they represent. Therefore Openness items, that may not necessarily be job-related, still offer insight into faking behavior. However, relying on items that do not result in disagreement due to item ambiguity alone may strengthen this approach due to the fact that responses to ambiguous items may change over time simply because participants simply do not know how to answer and do not remember what option they responded with on the previous occasion. Relying on items that represent more straightforward concepts that still result in changing scores and disagreement between conditions (if enough of such items exist to maintain functionality of the approach) may represent the ideal subset of items with which to construct the faking indicator score. Disagreement on these items would most likely represent differences in perception as to the most desirable response option, without contamination due to misunderstanding of the item(s) alone.

These results also suggest that using a more quantitative approach to the recoding scheme is preferable to relying on judgment alone. While the differences between the two recoding styles were often minimal, the quantitative method consistently performed at the same level or better than the qualitative approach for faking detection and overall performance, and often outperformed it or performed comparably in minimizing false-

positive decisions. Since the high-stakes world of hiring decisions depends on making accurate predictions and decisions, even small improvements are important. At the individual level, if one less honest responder is displaced due to faking or one less false-positive decision is made because of the use of the quantitative approach, this would represent a profoundly positive impact. Relatedly, while the quantitative method evidenced no correlations with honest condition personality scores, the qualitative method evidenced significant correlations for four of five personality factors. This suggests that faking (amongst real-world applicants) occurs in such a manner that differences between conditions may be minimized when viewing them judgmentally, yet become revealed when applying a more quantitative approach.

Implications for Research and Theory

Future research should attempt to assess this method of faking detection similarly with a within-subjects design, with less time between conditions that are counterbalanced for order effects, while using a larger sample of real-world job applicants from a more diverse array (still analyzed respectively) of job families and cultural backgrounds. Decreasing the time between conditions, or attempting to account for time effects with the implementation of assessment conditions that are counterbalanced for order effects, would be helpful in controlling for the possibility that individuals' scores have changed due to actual personality changes between assessments. Further, assessing the effectiveness of this method, both between and within respective cultures, may provide important information regarding its usefulness and potential limitations. Also, while it

may remain important to segregate job families at the time of analysis, establishing the utility of this approach for diverse occupations is necessary.

Additionally, further refinement of the recoding scheme, cut-scores and item selection method could be useful in increasing the accuracy of predictions and decreasing the occurrence of false-positive faking decisions, perhaps to the point that the quantitative indicator ultimately outperforms the qualitative approach in all three relevant phases (faking detection, avoiding false positive identification, and correct decision proportion). For instance, an even more granular recoding system may serve to increment the validity of the method with small differences between applicants compounding over multiple selected items, such that differential prediction occurs as a result. Analyzing at more numerous cut-scores (such as at $\frac{1}{4}$ or $\frac{1}{2}$ SD increments) might result in identifying the best possible combination of maximizing detection while minimizing false-positive decisions. Also, incorporating a highly trained panel of raters to assess the potential of each item for faking detection, and subsequently weighting the selected items according to their perceived potential and respective rater consensus could prove highly valuable in maximizing the potential of this approach.

Researchers should also attempt to incorporate individual differences measures while using real-world applicants. It may be that the low correlations with individual difference measures found in a directed-faking, lab-setting disappear when individuals are left to fake upon their own accord (Kuncel & Borneman, 2007). While correlations between this method and the research personality measures were found to remain low with my quantitative approach, they became significant for four of the five factors when

using the original qualitative approach. Further research into the effects of individual differences upon this method of faking detection should examine these relationships and the causes for the differences in personality correlations found here between the two approaches. Relatedly, the respective indicators were not correlated, suggesting that they may be detecting different types of fakers. Future research should investigate this further.

Further research should also be conducted to assess the relation between this method of faking detection and future work outcomes. Additionally, this should be done with multiple methods of true faking categorization. Do those individuals identified as faking job-related personality traits (by both the detection method and the type of categorization) evidence lower levels of criterion-related validity? Are there lower levels of performance and/or satisfaction and higher levels of turnover among these individuals? Relating this method of faking to criterion-related validity coefficients would go far in establishing the validity of this approach, as well as that of the various methods of true faking categorization.

Future researchers should also analyze the nature of this type of faking detection at the factor and facet level of the Big Five. It would be informative to understand whether certain factors or facets are more (or less) consistently identified as being faked using this approach, both within and between diverse occupations. Researchers should also expand this approach by analyzing personality score faking at the more granular facet-level. Does analyzing score changes at the facet-level impact the utility of this approach?

In addition, work should be done to determine if different combinations of the factors or facets represented by the items selected for use in comprising the faking indicator score affects the validity of this method. For instance, not including notoriously ambiguous Openness items for use in constructing the indicator score may improve the validity of this method, by somewhat controlling for the possibility that changes occur due to ambiguity, misinterpretation, or simply forgetting previous responses rather than intentional faking for such items.

Finally, previous research has suggested that work-contextualized measures of personality may result in increases in criterion-related validity coefficients (Shaffer, & Postlethwaite, 2012). Future research regarding this method should attempt to determine the impact of such measures on the implementation of this method of faking detection. It seems that standardizing the indicator scores should have served as a control for some of these effects. Comparing a contextualized measure that was recoded with unstandardized indicator scores, to a non-contextualized measure recoded with standardized indicator scores, would help researchers determine whether the theoretical notion of accounting for contextualization effects with standardized indicators is warranted.

CHAPTER VII

CONCLUSION

The previously studied methods for detecting or minimizing the occurrence of faking have mostly met with minimal success. The Kuncel and Borneman (2007) method to detecting faking represents a novel approach to the problem that has reported encouraging results. The current study's improvements, made through quantifying the recoding scheme and testing its efficacy with real-world applicants, a common personality measure, and a single job family, provide additional reason to remain positive about the potential utility of this method. With additional research and refinement of the underlying processes affecting the results found here, the application of this method may well represent the control for faking behavior researchers have sought after for so long.

REFERENCES

- Abrahams, N. M., Neumann, I., & Githens, W. H. (1971). Faking Vocational Interests: Simulated Versus Real Life Motivation 1. *Personnel Psychology*, 24(1), 5-12.
- Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The Magnitude and Extent of Cheating and Response Distortion Effects on Unproctored Internet-Based Tests of Cognitive Ability and Personality. *International Journal of Selection and Assessment*, 18(1), 1-16.
- Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology*, 17(1), 87-105.
- Austin, J. S. (1992). The detection of fake good and fake bad on the MMPI-2. *Educational and Psychological Measurement*, 52(3), 669-674.
- Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, 7(1), 84-92.
- Bagby, R. M., Gillis, J. R., & Dickens, S. (1990). Detection of dissimulation with the new generation of objective personality measures. *Behavioral Sciences & the Law*, 8(1), 93-102.
- Bagby, R. M., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of defensive responding in clinical and nonclinical samples. *Psychological Assessment*, 9(4), 406-413.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44(1), 1-26.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81(3), 261-272.
- Barrick, M. R., Mount, M. K., & 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(1-2), 9-30.

- Behling, O. (1998). Employee selection: will intelligence and conscientiousness do the job? *Academy of Management Executive* (1993-2005), 77-86.
- Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62(4), 833-863.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment*, 14(4), 317-335.
- Butcher, J. N., Morfitt, R. C., Rouse, S. V., & Holden, R. R. (1997). Reducing MMPI-2 defensiveness: The effect of specialized instructions on retest validity in a job applicant sample. *Journal of Personality Assessment*, 68(2), 385-401.
- Butcher, J. N., & Tellegen, A. (1966). Objections to MMPI items. *Journal of Consulting Psychology*, 30(6), 527-534.
- Campbell, J. P. (1990). An overview of the army selection and classification project (Project A). *Personnel Psychology*, 43(2), 231-239.
- Castro, S. L. (2002). Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, rwg (j), hierarchical linear modeling, within-and between-analysis, and random group resampling. *The Leadership Quarterly*, 13(1), 69-93.
- Cattell, H. E., & Mead, A. D. (2008). The sixteen personality factor questionnaire (16PF). *The SAGE handbook of personality theory and assessment*, 2, 135-159.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18(3), 267-307.
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for Faking: Effects on Criterion-Related Validity and Individual Hiring Decisions. *Personnel Psychology*, 47(4), 847-860.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press.
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, 16(2), 155-169.
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2006). Forcing choices in personality measurement. In R. L. Griffith, & M. H. Peterson, (Eds.), *A closer examination of applicant faking behavior* (pp.263-282). IAP.

- Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment*, 17(1), 47-60.
- Costa, P. T. (1996). Work and Personality: Use of the NEO-PI-R in Industrial/Organizational Psychology. *Applied Psychology*, 225-241.
- Costa Jr, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment*, 68(1), 86-94.
- Costa, P. T., & McCrae, R. R. personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual, 1992. *Psychological Assessment Resources*, Odessa, FL.
- Costa, P. T., McCrae, R. R., & Holland, J. L. (1984). Personality and vocational interests in an adult sample. *Journal of Applied Psychology*, 69(3), 390-400.
- Costa, P. T., McCrae, R. R., & Kay, G. G. (1995). Persons, places, and personality: Career assessment using the Revised NEO Personality Inventory. *Journal of Career Assessment*, 3(2), 123-139.
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, 42(1), 25-36.
- Denis, P. L., Morin, D., & Guindon, C. (2010). Exploring the Capacity of NEO PI-R Facets to Predict Job Performance in Two French-Canadian Samples. *International Journal of Selection and Assessment*, 18(2), 201-207.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440.
- Dilchert, S. & Ones, D. S. (2011). Application of preventive strategies. In M. Ziegler, C., & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 177-200). Oxford University Press.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, 16(1), 81-106.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40-57.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16(1), 1-23.

- Ellingson, J. E., Heggstad, E. D., & Makarius, E. E. (2012). Personality retesting for managing intentional distortion. *Journal of personality and social psychology*, 102(5), 1063-1076.
- Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: insights into response distortion. *Journal of Applied Psychology*, 92(2), 386-395.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84(2), 155-166.
- Erickson, P. B. (2004). Employer hiring tests grow sophisticated in quest for insight about applicants. *Knight Ridder Tribune Business News*, 1.
- Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, 97(4), 866-880.
- Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality*, 26(2), 103-120.
- Framingham, J. (2011). Minnesota Multiphasic Personality Inventory (MMPI). *Psych Central*. Retrieved on May 16, 2014, from <http://psychcentral.com/lib/minnesota-multiphasic-personality-inventory-mmapi/0005959>
- Furnham, A. F. (1997). Knowing and faking one's five-factor personality score. *Journal of Personality Assessment*, 69(1), 229-243.
- Gandy, J. A., Dye, D. A., & MacLane, C. N. (1994). Federal government selection: The individual achievement record.
- Ghiselli, E. E., & Barthol, R. P. (1953). The validity of personality inventories in the selection of employees. *Journal of Applied Psychology*, 37(1), 18-20.
- 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. L., & Peterson, M. H. (Eds.). (2006). *A closer examination of applicant faking behavior*. IAP.
- Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology*, 1(3), 308-311.

- Griffith, R. L., & Peterson, M. H. (2011). One piece at a time: the puzzle of applicant faking and a call for theory. *Human Performance*, 24(4), 291-301.
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36(3), 341-355.
- Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie canadienne*, 50(3), 151-160.
- Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, 11(4), 340-344.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26-42.
- Gonzalez-Mulé, E., Mount, M. K., Oh, I-S. (in press). A meta-analysis of the relationship between general mental ability and non-task performance. *Journal of Applied Psychology*.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135-164.
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91(1), 9-24.
- Heller, M. (2005). Court ruling that employer's integrity test violated ADA could open door to litigation. *Workforce Management*, 84(9), 74-77.
- Hills, P., & Argyle, M. (2001). Emotional stability as a major dimension of happiness. *Personality and Individual Differences*, 31(8), 1357-1364.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270-1285.
- Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied psychology*, 74(2), 273-279.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: a socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100.
- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, 18(4), 331-341.
- Hogan, R. T. (1991). Personality and personality measurement.

- Holden, R. R. & Book, A. S. (2011). Faking does distort self-report personality assessment. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 71-84). Oxford University Press.
- Holden, R. R., Fekken, G. C., & Cotton, D. H. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(1), 111-118.
- Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, 29(3), 362-372.
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63(2), 272-279.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources.
- Holland, J. L., Johnston, J. A., Hughey, K. F., & Asama, N. F. (1991). Some explorations of a theory of careers: VII. A replication and some possible extensions. *Journal of Career Development*, 18(2), 91-100.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11(2-3), 209-244.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75(5), 581-595.
- Hough, L. M. & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of Industrial, Work & Organizational Psychology: Volume 1: Personnel Psychology* (pp. 233-277). Sage.
- Hough, L. M., & Oswald, F. L. (2005). They're right, well... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, 18(4), 373-387.
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, 1(3), 272-290.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9(1-2), 152-194.

- Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, 53(2), 278-295.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72-98.
- Hurtz, G. M., & Alliger, G. M. (2002). Influence of coaching on integrity test performance and unlikely virtues scale scores. *Human Performance*, 15(3), 255-273.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: the Big Five revisited. *Journal of Applied Psychology*, 85(6), 869-879.
- Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2012). Emotional Intelligence in Personnel Selection: Applicant reactions, criterion, and incremental validity. *International Journal of Selection and Assessment*, 20(3), 347-358.
- Iliescu, D., Ilie, A., Ispas, D., & Ion, A. (2013). Examining the psychometric properties of the Mayer-Salovey-Caruso Emotional Intelligence Test: Findings from an Eastern European culture. *European Journal of Psychological Assessment*, 29(2), 121-128.
- Ispas, D., Iliescu, D., Ilie, A., & Johnson, R. E. (2014). Exploring the Cross-Cultural Generalizability of the Five-Factor Model of Personality The Romanian NEO PI-R. *Journal of Cross-Cultural Psychology*, 0022022114534769.
- Ispas, D., Iliescu, D., Ilie, A., Sulea, C., Askew, K., Rohlf, J. T., & Whalen, K. (2014). Revisiting the relationship between impression management and job performance. *Journal of Research in Personality*, 51, 47-53.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13(4), 371-388.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of applied psychology*, 69(1), 85-98.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Johnson, J. A., & Hogan, R. (2006). A socioanalytic view of faking. In R. L. Griffith, & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 209-231). IAP.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3), 621-652.

- Judge, T. A., Piccolo, R. F., & Kosalka, T. (2009). The bright and dark sides of leader traits: A review and theoretical extension of the leader trait paradigm. *The Leadership Quarterly*, 20(6), 855-875.
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives.
- Kaiser, R. B., & Hogan, J. (2011). Personality, leader behavior, and overdoing it. *Consulting Psychology Journal: Practice and Research*, 63(4), 219-242.
- Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, 93(1), 140-154.
- Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology*, 43(1), 48-55.
- 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(2), 220-231.
- Le, H., Oh, I. S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96(1), 113.
- Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies, and bullshit: Understanding faking and its effects on selection decisions using theory, field research, and simulation. In J. M. Brett & F. Drasgow (Eds.). *The psychology of work: Theoretically based empirical research* (pp. 253-276). Psychology Press.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14(2), 131-141.
- Li, N., Barrick, M. R., Zimmerman, R. D., & Chiaburu, D. S. (2014). Retaining the Productive Employee: The Role of Personality. *The Academy of Management Annals*, 8(1), 347-395.
- MacCann, C., Ziegler, M., & Roberts, R. (2011). Faking in personality assessments: reflections and recommendations. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 309-329). Oxford University Press.
- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, 32(2), 247-256.

- McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509-516.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: more substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882-888.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90.
- McCrae, R. R., Costa, P. T., Del Pilar, G. H., Rolland, J. P., & Parker, W. D. (1998). Cross-Cultural Assessment of the Five-Factor Model The Revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, 29(1), 171-188.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85(5), 812-821.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43(2), 335-354.
- Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review of research designs and analytic strategies. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 85-113). IAP.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). 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.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88(2), 348-355.
- Murphy, K. R. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18(4), 343-357.
- Newman, D. A., & Lyon, J. S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology*, 94(2), 298-317.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60(4), 995-1027.

- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human performance*, 11(2-3), 245-269.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81(6), 660-679.
- Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, 14(2), 150-163.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.
- Peterson, M. H., Griffith, R. L., & Converse, P. D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single-and multiple-predictor selection. *Journal of Business and Psychology*, 24(4), 373-386.
- Peterson, M. H., Griffith, R. L., Converse, P. D., & Gammon, A. R. (2011). Using within-subjects designs to detect applicant faking. In *26th Annual Conference for the Society for Industrial/Organizational Psychology*, Chicago, IL.
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78(3), 582-593.
- Piedmont, R. L., & Weinstein, H. P. (1994). Predicting supervisor ratings of job performance using the NEO Personality Inventory. *The Journal of Psychology*, 128(3), 255-265.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153-172.
- Popham, S. M., & Holden, R. R. (1990). Assessing MMPI constructs through the measurement of response latencies. *Journal of Personality Assessment*, 54(3-4), 469-478.
- Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment*, 13(4), 304-315.
- 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(3), 241-258.

- Raymark, P. H., & Tafero, T. L. (2009). Individual differences in the ability to fake on personality measures. *Human Performance*, 22(1), 86-103.
- Reeder, M. C. & Ryan, A. M. (2011). Methods for correcting faking. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 131-150). Oxford University Press.
- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 2(1), 5-6.
- Robie, C., Curtin, P. J., Foster, T. C., Phillips IV, H. L., Zbylut, M., & Tetrick, L. E. (2000). The effects of coaching on the utility of response latencies in detecting fakers on a personality measure. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 32(4), 226-233.
- Rosse, J. G., Levin, R. A., & Nowicki, M. D. (1999). Assessing the impact of faking on job performance and counter-productive job behaviors. In P. Sackett (Chair), *New empirical research on social desirability in personality measurement. Symposium conducted at the 14th annual meeting of the Society of Industrial Organizational Psychology, Atlanta, GA*.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634-644.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, 16(2), 155-180.
- Ryan, A. M., Ployhart, R. E., & Friedel, L. A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology*, 83(2), 298-307.
- Salgado, J. F. (1997). The Five Factor Model of personality and job performance in the European Community. *Journal of Applied psychology*, 82(1), 30-43.
- Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, 11(2-3), 271-288.
- Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, 43(1), 133-168.
- 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(2), 262-274.

- Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, 91(3), 613.
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A Matter of Context: A Meta-Analytic Investigation of the Relative Validity of Contextualized and Noncontextualized Personality Measures. *Personnel Psychology*, 65(3), 445-494.
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87(2), 211-219.
- Smith, D. B., & McDaniel, M. (2011). Questioning old assumptions: Faking and the personality-performance relationship. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 53-70). Oxford University Press.
- Smith, D. B., & Robie, C. (2004). The implications of impression management for personality research in organizations. *Personality and Organizations*, 111-138.
- Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9(2), 219-242.
- Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86(5), 943-953.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology*, 89(3), 497-508.
- Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. *The SAGE handbook of personality theory and assessment*, 2, 261-292.
- Tett, R. P., Anderson, M. G., Ho, C., Yang, T. S., Huang, L., & Hanvongse, A. (2006). Seven nested questions about faking on personality tests: An overview and interactionist model of item-level response distortion. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior*, 43-83.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60(4), 967-993.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology*, 44(4), 703-742.

- Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, 23(1), 117-124.
- Uziel, L. (2010). Rethinking Social Desirability Scales From Impression Management to Interpersonally Oriented Self-Control. *Perspectives on Psychological Science*, 5(3), 243-262.
- 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.
- Vispoel, W. P., & Tao, S. (2013). A generalizability analysis of score consistency for the Balanced Inventory of Desirable Responding. *Psychological Assessment*, 25(1), 94-104.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197-210.
- Winkelspecht, C., Lewis, P., & Thomas, A. (2006). Potential effects of faking on the NEO-PI-R: Willingness and ability to fake changes who gets hired in simulated selection decisions. *Journal of Business and Psychology*, 21(2), 243-259.
- Wonderlic Personnel Test (1992) Wonderlic Personnel Test & Scholastic Level Exam User's Manual. Milwaukee, WI: Author.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20(1), 71-87.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84(4), 551-563.
- Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3-16). Oxford University Press.

APPENDIX A
DECOMPOSITION OF TRUE FAKING
CATEGORIZATION METHODS

Aside from the two methods that were decided upon for true faking categorization (and previously detailed in the method section), six other methods for making such categorizations were examined. I have included the details of all these methods here, repeating those of the two that were previously outlined to facilitate reference between respective methods.

SEM (1 CI) used one 95% CI built around the scores in the honest condition. Following the formula used in Hogan et al. (2007), SEM was calculated by multiplying the SD of the research condition scores by the square root of the quantity of one minus the squared reliability [$\sigma * \sqrt{(1 - r^2)}$]. The 95% confidence interval was then established by multiplying the resulting value by 1.96. For the respective personality factors, if a participant's scores in the applicant condition fell outside of their scores from the research condition +/- the value calculated for the 95% CI using the SEM, then that applicant was categorized as a faker. Regarding SEM (1 CI) for Conscientiousness, approximately 5% (11/213) of the sample was found to have an applicant score that exceeded these limits and was subsequently categorized as true fakers. For Neuroticism, approximately 5% (10/213) of the sample was also found to have an applicant score that

exceeded these limits. For Extraversion, less than 1% (1/213) of the sample was found to have an applicant score that exceeded these limits.

SEM (2 CI) used two 95% CI's; one built around the honest scores and one around the faked scores. These CI's were calculated in the same manner in which the CI was calculated for the SEM (1 CI) method, with the exception that the CI for the faking scores was calculated using the reliability and SD for the scores from the faking condition. For the respective personality factors, if an applicant's CI from the research condition to the applicant condition did not overlap, then that applicant was categorized as a faker. Regarding the SEM (2 CI) approach, no individuals (0/213) in the sample were found to have CI's that did not overlap and were subsequently labeled true fakers for any of the three predictors.

Following the method used in Griffith et al. (2007), SED was calculated by multiplying SEM by 1.4, which results in a more conservative CI and identifies more extreme fakers. From there, the SED (1CI) and SED (2 CI) methods were conducted identically to the corresponding methods (using the SEM) that were previously discussed. Regarding SED (1 CI) for Conscientiousness, approximately 2% (4/213) of the sample was found to have an applicant score that exceeded these limits, and was subsequently labeled true fakers. For Neuroticism, less than 1% (1/213) of the sample was found to have an applicant score that exceeded these limits. For Extraversion, less than 1% (1/213) of the sample was also found to have an applicant score that exceeded these limits. Regarding the SED (2 CI) approach, no individuals (0/213) in the sample were

found to have CI's that did not overlap and were subsequently labeled true fakers for any of the three respective predictors.

Following the formula used in Arthur et al. (2010), SEM_d was calculated by multiplying the SD of the difference scores (between research and applicant conditions) by the square root of the quantity of one minus the squared research/applicant correlation [$\sigma * \sqrt{(1 - r_{12}^2)}$]. For the respective personality factors, if an applicant's change score was greater than the absolute value of SEM_d , that applicant was categorized as a faker. For Conscientiousness, approximately 69% (146/213) of the sample was found to have exceeded this limit with their change in scores and were subsequently labeled true fakers. For Neuroticism, approximately 54% (114/213) of the sample was found to have either raised or lowered their scores beyond this limit. For Extraversion, approximately 46% (99/213) of the sample was found to have either raised or lowered their scores beyond this limit.

McFarland and Ryan's (2000) formula to calculate the reliability of change scores (research/applicant) was calculated as well. This was done following the Hogan et al. (2007) approach that calculated SEM for the difference scores in an attempt to make faking categorizations. The rationale behind such a calculation is similar to that of the SEM_d procedure above, although it uses a different formula. The reliability of change scores here was calculated in two steps. First, by multiplying the variance for the applicant and research conditions respectively by the quantity of one minus their corresponding reliabilities, then summing these resulting values [$\sigma_a^2(1-r_a) + \sigma_r^2(1-r_r)$]. Then, the quantity of this value subtracted from the variance of the change scores was

divided by the variance of the difference scores $[(\sigma_d^2 - [\sigma_a^2(1-r_a) + \sigma_r^2(1-r_r)]) / \sigma_d^2]$.

However, conducting these calculations resulted in negative reliabilities for the change scores. An examination of these results revealed variances (from the current study's sample) for the factor scales that were much greater than those from the study in which this formula was developed. These high variances were the cause of the change score reliability calculations resulting in negative (and therefore unusable) values.

The $> \pm 1SD + |M \text{ Change}|$ method used the mean difference (M_D) between research condition scores and application condition scores for Conscientiousness ($M = 6.41$, $SD = 7.95$), Neuroticism ($M = -3.35$, $SD = 7.87$), and Extraversion ($M = 2.25$, $SD = 7.44$). The absolute value of the sum of the SD of the difference scores and the M_D , resulted in a threshold of ± 14.43 for change in Conscientiousness scores, ± 11.22 for Neuroticism scores, and ± 9.69 for Extraversion scores. Change in either direction beyond these respective thresholds resulted in a true faking categorization. For Conscientiousness, approximately 13% (28/213) of the sample was found to have exceeded this limit with their change in scores and were subsequently labeled true fakers. For Neuroticism, approximately 15% (33/213) of the sample was found to have either raised or lowered their scores beyond this limit. For Extraversion, approximately 25% (53/213) of the sample was found to have either raised or lowered their scores beyond this limit.

The $> \pm \frac{1}{2} SD \text{ Change}$ method used thresholds determined by the observed SD from the honest condition. If participants changed their scores in the faking condition by more than $\frac{1}{2} SD$ (honest condition), then those participants were labeled as fakers. For

Conscientiousness ($SD = 20.15$), this resulted in a threshold of ± 10.07 with approximately 31% (67/213) of the sample found to have either raised or lowered their scores beyond this limit and subsequently labeled true fakers. For Neuroticism ($SD = 20.83$), this resulted in a threshold of 10.42 with approximately 20% (42/213) of the sample found to have either raised or lowered their scores beyond this limit. For Extraversion ($SD = 18.40$), this resulted in a threshold of 9.20 with approximately 25% (53/213) of the sample found to have either raised or lowered their scores beyond this limit.

APPENDIX B

FIGURES DEPICTING COMPARISONS
OF THE RESPECTIVE METHODS

Figure 5. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) for the Entire Sample for the Respective Predictors.

Figure 6. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made for the Entire Sample.

Figure 7. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion for the Entire Sample.

Figure 8. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) for the Entire Sample for the Respective Predictors.

Figure 9. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made for the Entire Sample.

Figure 10. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion for the Entire Sample.

Figure 11. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 12. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 13. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 14. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 15. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 16. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Conscientiousness Scores.

Figure 17. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Neuroticism Scores.

Figure 18. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Neuroticism Scores.

Figure 19. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Neuroticism Scores.

Figure 20. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Neuroticism Scores.

Figure 21. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Neuroticism Scores.

Figure 22. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Neuroticism Scores.

Figure 23. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Extraversion Scores.

Figure 24. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Extraversion Scores.

Figure 25. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Extraversion Scores.

Figure 26. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Three Respective Selection Percentages for Extraversion Scores.

Figure 27. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Three Respective Selection Percentages for Extraversion Scores.

Figure 28. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Three Respective Selection Percentages for Extraversion Scores.

Figure 29. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 30. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 31. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 32. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 33. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 34. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Conscientiousness Scores.

Figure 35. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 36. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 37. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 38. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 39. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 40. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Neuroticism Scores.

Figure 41. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 42. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 43. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 44. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 45. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 46. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion at Two Respective Select-Out Thresholds for Extraversion Scores.

Figure 47. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) after Removing the Top and Bottom 10% for Three Predictors.

Figure 48. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made after Removing the Top and Bottom 10% for Three Predictors.

Figure 49. Comparison of the Quantitative and Qualitative Methods of Detection Using the 1 SD Method of True Faking Categorization and the Correct Decision Proportion after Removing the Top and Bottom 10% for Three Predictors.

Figure 50. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Percentage of Fakers Identified (Relative to Those Present) after Removing the Top and Bottom 10% for Three Predictors.

Figure 51. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Number of False-Positive Faking Identifications Made after Removing the Top and Bottom 10% for Three Predictors.

Figure 52. Comparison of the Quantitative and Qualitative Methods of Detection Using the $\frac{1}{2}$ SD Method of True Faking Categorization and the Correct Decision Proportion after Removing the Top and Bottom 10% for Three Predictors.