2013

Unproctored internet testing: Methods to mitigate and detect cheating

Emily Annette Shindledecker

UNPROCTORED INTERNET TESTING:

METHODS TO MITIGATE AND DETECT CHEATING

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Psychology:

Industrial/Organizational

---

by

Emily Annette Shindledecker

December 2013

UNPROCTORED INTERNET TESTING:

METHODS TO MITIGATE AND DETECT CHEATING

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Emily Annette Shindledecker

December 2013

Approved by:

[redacted]
Dr. Kenneth Shultz, Chair, Psychology

11/8/13
Date

[redacted]
Dr. Janet Kottke

[redacted]
Dr. Matt Riggs

ABSTRACT

Unproctored internet testing (UIT) is becoming increasingly prevalent for personnel selection. However, the issues and implications of cheating in such tests are largely unknown. Therefore, the purpose of this study was to examine cheating in UIT settings. Cognitive ability test scores and personality measures were collected through convenience sampling methods and a total of 147 participants provided useable responses. A 2x2 between subjects factorial analysis of variance (ANOVA) was conducted to assess significant differences between stakes and instruction conditions on cheating. A simultaneous multiple regression was conducted to analyze personality characteristics (i.e., honesty, narcissism, Machiavellianism, and psychopathy) as predictors of cheating. A 2x2 ANCOVA was conducted to assess the impact of stakes and instruction conditions on cheating after controlling for personality characteristics. Results showed no main effects for stakes or instruction conditions on cheating; personality characteristics did not significantly predict cheating; and the impact of stakes and instruction conditions were non-significant after controlling for personality characteristics. Implications for practice

suggest that organizations not only need to be mindful when deciding to administer UIT for employee selection, but also determine if a cheating detection formula will be utilized and if this method will be legally defensible.

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Kenneth Shultz, for his continued investment in this study.

I would also like to thank the Industrial and Organizational faculty for assisting in my professional growth and development.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

Selection in Organizations

Organizations are becoming increasingly aware that attracting and retaining talented employees will provide a competitive advantage (Carless, 2009). Specifically, studies have indicated that organizations that select best fitting candidates are likely to benefit from increased profitability (Huselid 1995; Terpstra & Rozell, 1993). Therefore, selecting talented employees leads to competition among organizations for top candidates. In order to remain at a competitive edge, numerous organizations are beginning to adopt best selection practices. Among the best selection practices is psychological testing, which has been regarded as an essential component to the selection process for more than a century (Guest, Michie, Conway, & Sheehan, 2003; Harel & Tzafrir, 2001; Terpstra & Rozell, 1993).

The use of psychological testing for selection purposes is on the rise (Anderson, 2005; Bartram, 2001; Taylor, Keelty & McDonnell, 2002; Wolf & Jenkins, 2006). There are a variety of reasons for the increasing use of

psychological testing (Carless, 2009). First, psychological tests are objective, and therefore, reduce the prospect for legal challenges, as job candidates often perceive these as fairer compared to less objective selection devices such as unstructured interviews and role plays (Carless, 2009). Second, numerous psychological tests have predictive validity and reliability, allowing for greater accuracy when predicting future job performance (Carless, 2009). Finally, with the widespread use and availability of the internet, psychological tests may be administered online, resulting in a decrease of costs for organizations, thus increasing organization's return on investment (ROI) (Carless, 2009).

The use of technology in the 21st century is increasing at a rapid pace, especially in the workplace. Specifically, advances in technology have assisted in the shift from paper-and-pencil tests to computer-based tests used in employee selection (Makransky & Glas, 2011). With the widespread use of computer-based assessments, organizations are beginning to implement unproctored internet testing (UIT). Although unproctored internet testing offers numerous advantages over proctored testing (i.e., reduced cost, administrative ease, and flexibility), researchers

and practitioners are concerned with the increased ease of cheating and response distortion. Additionally, Beaty, Nye, Borneman, Kantrowitz, Drasgow, and Grauer (2011) state that unproctored internet cheating is still a relatively unexplored topic in the literature and future research would be highly beneficial for the field of human resource selection, both in terms of advancing research as well as applied practice.

The current research on cheating in unproctored internet testing is timely to industrial/organizational psychology, as it is vital for organizations to determine which conditions candidates are most likely to engage in cheating as well as how to best mitigate cheating in assessments. The intention of this proposed study will be to investigate the effects of high versus low-stakes conditions and the presence or absence of instructions on cheating in UIT settings. This literature review will provide a comprehensive description of UIT; the advantages and disadvantages of UIT; malfeasant behaviors; types of cheating; cheating conditions; methods to mitigate cheating; the statistical z-test to detect cheating; seek to identify characteristics of cheaters; state methods

performed; and include practical implications for future directions in research and practice.

Unproctored Internet Testing

Unproctored internet testing (UIT) is defined as "a testing event that is not monitored by a human test administrator. Consequently, candidate identification is not verified by a human proctor nor is candidate behavior during the testing event observed by a human" (Tippins, et al., 2006, p. 193). In a recent study, Fallaw, Solomonson, and McClelland (2009) found that over two thirds of employers surveyed are currently engaging in UIT for employee selection purposes. Therefore, with numerous organizations currently implementing UIT, research needs to focus on strategies to enhance this method for employee selection purposes, especially considering that the practice of UIT has far outpaced the science behind UIT (Pearlman, 2009).

Unproctored Internet Testing: Advantages

There are a variety of advantages and disadvantages when utilizing UIT as a human resource selection method. Specifically, the advantages focus on improving the testing process by reducing costs and maximizing efficiency (Nye,

Do, Drasgow, & Fine, 2008). The first primary advantage of UIT includes reducing the costs associated with selection procedures. This is accomplished through screening a large pool of applicants and moving only qualified applicants on to the next hurdle in the selection process (Nye et al., 2008). Large organizations have reported receiving "tens of thousands applications per month," and with the increasing use of online job applications, efficiently and effectively reducing the number of initial applicants is key to cost savings (Nye et al, 2008, p. 113).

The second advantage of UIT is the idea that candidates may be assessed from virtually any geographic location with access to the internet. This allows for an increase in the overall applicant pool requiring candidates to travel to a physical location to take the exam as well as a potential increase in applicant diversity (Nye et al., 2008). By increasing the overall applicant pool as well as the diversity, this allows for greater opportunity for organizations to select the best fitting candidates for the position(s).

The third advantage of UIT is that it allows for accommodation for individuals with special needs and/or disabilities. UIT provides an opportunity for individuals

with special needs and/or disabilities to be able to take the test in the convenience of their own home rather than struggling to meet at a physical location to take an assessment (Nye et al., 2008). However, if candidates require assistive technology, organizations may want to provide reasonable accommodations at a specified testing location.

The last advantage of UIT is that results are often available faster than traditional paper-and-pencil exams. This allows for a diminution in the time-to-hire ratio, which helps ensure that selected candidates are less likely to have not already pre-selected themselves out based on other job opportunities they may have received prior (Nye et al., 2008). Thus, the efficiency of time offered is often the most important factor when organizations choose to implement UIT in their selection process.

Unproctored Internet Testing: Disadvantages

UIT has a variety of attractive advantages to offer to organizations, however, it does not come without its disadvantages. The first primary disadvantage of UIT is that this process is limited by socio-economic-status (Nye et al., 2008). Therefore, individuals who do not have access to computers and the internet are negatively

impacted by the use of UIT. Specifically, results show that only 43% of lower socio-economic-status homes versus 93% of homes in higher socio-economic-status are connected to the internet (Pew Internet and American Life Project, 2007). In addition, when considering the worldwide internet usage rate of only 15.7%, this severely places the majority of individuals at a disadvantage of engaging in the UIT selection method (Naglieri, Drasgow, Schmite, Handler, Prifitera, Margolis, & Velasquez, 2004). If candidates do not have access to a computer or the internet, the organization must take responsibility for providing test-taking alternatives, such as administering the test at a specified location.

The second disadvantage of UIT is that individuals have a varying range of experience with the internet (Nye et al., 2008). Therefore, individuals with limited internet experience are at a disadvantage as compared to those with more advanced experience. It is also likely that performance on the test will be contaminated by technological ability when trying to measure cognitive ability or personality constructs (Nye et al., 2008). The third disadvantage of UIT is the initial costs of implementing new technology in the workplace. Nye et al.

(2008) has reported initial test development and implementation for various organizations has ranged in cost from $50,000 to three million dollars. However, it must be noted that often times the long-term cost savings outweigh the initial costs of development and implementation of new selection devices used in UIT.

The final, and most salient, disadvantage of UIT is susceptibility to malfeasant behaviors of candidates (Nye et al., 2008). It is extremely easy for candidates in UIT situations to engage in cheating behaviors, which makes the validity as well as candidate's true performance on the test difficult to assess. Based upon the pernicious nature of cheating and the need for research to further explore conditions in which cheating occurs as well as how to mitigate this issue, this proposed research will specifically address cheating in UIT.

Malfeasant Behaviors

Malfeasant behaviors have been defined in the literature as "deliberately falsifying one's responses on a test in an attempt to distort one's standing on the construct of interest" (Arthur, Glaze, Villado, & Taylor, 2010, p. 2). Specifically, malfeasant behaviors with regard

to employment testing can be broken down into two separate distinct constructs: response distortion and cheating.

Response Distortion

Response distortion refers to engaging in faking, impression management, and socially desirable responding (Arthur et al., 2010). Response distortion is associated with personality and non-cognitive measures based upon the lack of one correct answer and the perceived desired responses of an organization. Various studies report a lack of differences in scores of personality and non-cognitive measures between unproctored and proctored settings (Arthur et al, 2009; Arthur et al. 2010; Nye et al., 2008; & Tippins et al., 2006). The presence of a proctor is unlikely to stop candidates from responding in socially desirable fashion (Nye et al., 2008). The preexisting research on response distortion in non-cognitive and personality measures and the unchanging nature of scores between unproctored and proctored test settings, reduce the need to further research response distortion in this particular study. Therefore, the present study will center on cheating employed in employment testing.

Cheating

Cheating refers to utilizing outside sources (e.g., dictionaries, calculators, Google internet searches, etc.), obtaining information about test questions prior to the assessment, and/or utilizing surrogate test-takers (e.g., a friend or family member) (Arthur et al., 2010). Cheating is associated with cognitive ability and knowledge-based tests. These tests require that participants must choose the one "correct" answer, and are often designed to measure maximal performance. Cognitive ability tests have been supported as most predictive of job performance, which makes this predictor an extremely important construct of interest to study (Guo, Tay, & Drasgow, 2009). Whitley (1998) and Cizek (1999) have reported that approximately half of all college students have admitted to cheating on tests during the course of their college careers. Additionally, Cizek (1999) reports that 45% of candidates have admitted to falsifying work experience AND would be willing to cheat on a selection test if provided the opportunity. The above results demonstrate the high likelihood of cheating, and the probability of cheating is expected to be higher in unproctored settings (Makransky & Glas, 2011). Survey research has found that cheating is the

number one reason that human resources (HR) professionals are resistant to implementing UIT (Pearlman, 2009). Therefore, in order to allow HR professionals more comfort when utilizing unproctored internet tests, additional research needs to be conducted on conditions in which applicants feel most inclined to cheat and how to best mitigate this behavior- both of which were directly addressed in the present study.

Large-Scale Cheating

The literature has indicated two types of cheating: large-scale cheating and small-scale cheating. Large-scale cheating occurs when an entire organization (e.g., a testing company) collects and distributes items to test-takers from their "testing program" for a rather costly fee (Drasgow, 2009). This type of cheating has been found in examples such as the Kaplan incident of 1994. In this specific incident of large-scale cheating, Kaplan employees were sent to take the GRE examination. Subsequently, the employees memorized items from the examination, and developed a large testing pool to teach students to the test (Honan, 1995). This scandal resulted in a decline in Kaplan's reputation as well as an extremely costly lawsuit.

More recently, a large-scale cheating scandal occurred on the SAT in Long Island, New York (Anderson, 2012). This specific incident involved approximately 20 students; 5 students who served as surrogate test-takers, and 15 students who paid surrogates between $500 and $3,600 to take their standardized tests (Anderson, 2012). This scandal resulted in more stringent test security in which students are required to send in or upload their photo on their SAT or ACT admission ticket, as well as provide matching identification at the start of their exams. In addition, the SAT and ACT scores are now required to be sent to the students' high school, to ensure added post-test security (Anderson, 2012). However, based upon the difficulty in replicating a large-scale cheating scenario in an experimental setting as well as ethical issues that may arise, this will not be the focus of the present study. Rather, we will focus on small-scale cheating in UIT settings.

Small-Scale Cheating

In contrast, small-scale cheating occurs when individuals (e.g., conspirators) memorize test items and informs others (e.g., the beneficiaries) about the test questions in advance. Individuals may also utilize outside

testing sources to find correct answers, or surrogates are utilized in place of the test takers (Guo et al., 2009). Specifically, the effects of small-scale cheating inflates cut-points, which may result in disqualification of a large number of qualified applicants, who otherwise would have moved on to the next hurdle in the selection process (Pearlman, 2009). Additionally, small-scale cheating may negatively affect the qualified applicant pool, which is likely to increase turnover rates and be extremely costly for the organization long-term (Pearlman, 2009). Although there are a variety of small-scale cheating methods that result in negative outcomes for organizations, how applicants perceive the testing setting is likely to impact their decision to engage in cheating behaviors. Specifically, high-stakes and low-stakes testing will be discussed as follows.

Cheating in High-Stakes Versus Low-Stakes Conditions

High-stakes testing has been defined in the literature as test scores being utilized for selection, promotion decisions, and/or any other decision in which a candidate perceives important (Arthur et al., 2010). Low-stakes testing, on the other hand, has been defined as test scores being utilized for developmental or researching purposes

and/or the candidate does not feel as though much is at risk (Arthur et al., 2010).

In a study conducted by Arthur et al. (2010), participants were predicted to engage in higher levels of cheating in high stakes conditions versus low stakes conditions, based upon a highly valued outcome (e.g., promotion and grades). Specifically, results were expected to show that test scores from individuals in high-stakes conditions would have higher test scores than individuals in the low-stakes conditions (Arthur et al., 2010). Additionally, individuals in high-stakes conditions were predicted to have higher UIT scores than verification scores on a proctored exam, indicating that cheating was involved. Despite predictions, results showed that participants who took a speeded cognitive ability test in high-stakes condition actually received similar scores to the low-stakes condition (Arthur et al., 2010). Contrary to expectations, participants actually received higher scores on their follow-up proctored exam than in their pre-screened UIT. Results showed that only 7.7% of individuals received significantly lower scores (1 standard error of measurement lower) in their retest as opposed to their UIT (Arthur et al., 2010). Results of this study were justified

by explaining that retest scores were higher based upon practice effects that may have occurred. Thus, it will be beneficial for research to focus on cheating detection in high-stakes versus low-stakes conditions when controlling for practice effects.

Methods to Mitigate Cheating

The literature suggests a variety of methods to help organizations mitigate small-scale cheating, including: verification testing, providing instructions, and utilization of computer adaptive testing (Pearlman, 2009; Hense, Golden, & Burnett 2009; Guo et al., 2009). These three methods will be discussed in greater detail as follows.

Verification Testing

A commonly proposed and applied method to mitigate cheating in employee selection assessments is verification testing (Tippins et al., 2006). Verification testing involves an unproctored examination used as a pre-screen device, followed by a proctored administration used as a verification device (Pearlman, 2009). When individuals are required to take follow-up verification assessments, job applicants have been predicted to be less likely to engage

in cheating, but this has yet to be verified in the literature (Do, 2009).

It is important to note that verification testing does not come without its limitations. The literature suggests that implementation of verification testing increases steps in the recruitment process, which makes the assessment process more time-consuming and costly (Pearlman, 2009). However, verification testing is the most effective method of mitigating cheaters to date, and must continue to be researched on its overall effectiveness.

Instructions

In an article written by Hense et al. (2009), the authors state that a method to mitigate cheating is providing instructions to test-takers. This includes encouraging honest responding, and is based upon the premise that increased accountability will serve to discourage cheating intentions (Lievens & Burke, 2010). Additionally, Hense et al. (2009) utilized instructions to inform candidates that responses may be verified through an additional assessment and/or a structured interview with the hiring manager(s). The candidates were also instructed that if caught cheating, they would immediately be screened out of the selection process (Hense et al., 2009). In this

particular study, implementing instructions prior to assessment resulted in higher proctored retest scores as compared with the original unproctored retest scores, which is indicative of a lack of cheating as well as improvements due to practice effects (Hense et al., 2009). However, an important limitation of the Hense et al. (2009) study is that there was no control group. It would be highly beneficial to determine if a main effect exists for instruction vs. non-instruction conditions on cheating in UIT.

Computer Adaptive Testing

It is noted in the literature that small-scale cheating is highly prevalent to conventional forms of testing (Guo et al., 2009). Specifically, conventional forms of testing occur when the same test format and same test questions are continuously utilized (Guo et al., 2009). Thus, individuals have a greater opportunity to share answers with others prior to their test administration. Additionally, applicants taking a verification test may have inflated scores due to practice effects, which results from the same questions utilized. It is important to note that conventional test taking may be improved by offering parallel forms of differing test

questions and measuring parallel constructs. However, small-scale cheating has been found in the literature to be less prevalent when using computer adaptive testing (CAT) than conventional forms of testing (Guo et al., 2009). This is based upon the idea that CAT establishes a unique testing setting for each test-taker, which enhances test-security (Kantrowitz & Dawson, 2011). CAT often employs a timer for each question, thus minimizing the amount of time test-takers have to research answers through a dictionary or via internet (Kantrowitz & Dawson, 2011). Finally, CAT often includes a two-stage testing process. In the first stage, the unproctored internet test is administered. In the second stage, the organization administers a follow-up proctored verification test to confirm the first test score (Fetzer & Grelle, 2010).

Computer Adaptive Testing (CAT) is an emerging method that assists to mitigate against the challenges of cheating in UIT (Kantrowitz & Dawson, 2011). CAT is based on the foundation of a mathematical algorithm using item response theory (IRT), and test questions are chosen from large pools in order to adapt to each test-taker (Kantrowitz et al., 2011). Specifically, the initial test-item is selected for the test-takers. If the test-taker answers the item

correctly, he/she will receive a more difficult test question. However, if the test-taker answers an item incorrectly he/she will receive an easier test question (Guo et al., 2009). The test continues in this format until the computer can accurately provide a reliable assessment of the test taker's ability, or the pre-selected number of test items has been exhausted (Fetzer, Dainis, Lambert, & Meade, 2011). There are a variety of advantages and disadvantages when utilizing CAT in UIT settings.

Computer Adaptive Testing: Advantages

CAT provides a variety of attractive advantages for organizations. First, "CAT is a method of testing that 'adapts' to each individual test taker. In other words, CAT provides a tailored testing experience based on the test taker's level of knowledge, skill, ability, or other characteristic being evaluated by the test" (Fetzer et al., 2011, p. 2). Specifically, each CAT test that participants take, whether it be in an unproctored internet test or on a verification test, will be unique, which will help mitigate against practice effects often found on repeat measure conventional tests. Secondly, not only does the CAT adapt itself to test takers ability-level, but it also provides maximal efficiency and accuracy (Fetzer et al., 2011).

Fewer items may be used as well as provide a more accurate score for test takers than conventional forms of testing. This allows for a reduction in participant fatigue that may occur on longer conventional tests (Fetzer et al., 2011). Third, utilizing a CAT maximizes test security (Fetzer et al., 2011). Specifically, CAT is administered electronically, which eliminates the need for hard copies of the test. Therefore, test takers do not have the opportunity to distribute the test to others. Additionally, CAT provides varying test items, which decreases discussion about the test content among test-takers (Fetzer et al., 2011).

Computer Adaptive Testing: Disadvantages

Although the CAT offers a variety of attractive advantages, this testing format does not come without limitations. First, the CAT is extremely time consuming to develop and validate, particularly given such a large pool of test questions is required. Secondly, not only is the CAT time consuming, but extremely costly for organizations as well, which makes this type of test unfeasible for many organizations, especially small-businesses (Guo et al., 2009). The literature suggests that more and more organizations are beginning to utilize the CAT, and

therefore in order to stay ahead of practice, this study will focus on the use of the CAT over conventional forms of testing (Kantrowitz, 2011). There are a variety of methods that may be utilized to mitigate against cheating in UIT. However, if these methods are unsuccessful, organizations still need a method to detect cheating in UIT.

Cheating Detection

The Z-Test

In a study conducted by Guo and Drasgow (2010), the Z-test was identified as an "efficient and effective" method to detect cheating in UIT (p. 351). In order to perform the Z-test to detect cheating, three conditions must be met (Guo & Drasgow, 2010). First, an unproctored as well as a verification test must be provided to participants, and both tests need to be based upon item response theory. This condition may be met by providing participants with a computer adaptive test. Second, the unproctored and verification tests must both measure the same construct. This condition may be met by providing participants with a cognitive ability test in both UIT and verification testing. Third, the UIT and the verification test must be

unidimensional, which may be met by choosing a single construct of interest.

If the above three conditions are met, the Z-test may be utilized to detect cheating. The Z-test is based upon the premise of comparing ability estimates (Guo et al., 2010). If cheating did not occur, we are likely to see that the ability estimate of the UIT will equal the ability estimate of the verification test (ability estimates are indicated by the symbol theta). If cheating does occur, we are likely to see that the UIT ability estimate will not equal the ability estimate of the verification test. Specifically, if cheating occurs in the UIT we are likely to see that the ability estimate of the UIT will be greater than the ability estimate of the verification test. We may assume that the ability estimates of both the UIT and the verification test will be distributed normally, and thus, the difference between the two scores will approximate a normal distribution. Therefore, by calculating a standardized score, the z-test may be performed using the following formula (Guo et al., 2010):

$$Z = \frac{(\hat{\theta}_u - \hat{\theta}_v)}{\sqrt{SE_u^2 + SE_v^2}}$$

This formula calculates a standardized score for each participant by taking the ability estimate from the verification test, and subtracting this from the ability estimate from the unproctored test session (numerator). Furthermore, the square root is taken of the sum of the standard error of the estimate squared from the verification test and the standard error of the estimate squared from the unproctored test session (denominator). This formula allows for each test-taker to be compared in terms of cheating along a continuum.

We can assume that cheating is likely to occur on the UIT rather than the proctored test. This allows for a one-tailed test to be performed with Type 1 error alpha= .05. For practical application, if the z-statistic is > 1.645, participants will be classified as engaging in cheating behaviors on the UIT (Guo et al., 2010).

Characteristics of Cheaters

In the following sections, the Five-Factor model of personality, the HEXACO model of personality, and the dark triad constructs of personality will be discussed.

The Five-Factor Model of Personality

Nathanson, Paulhus, and Williams (2006) state that a great deal of research has been conducted on personality predictors of cheating. Specifically, a meta-analysis conducted by Whitley and Keith-Spiegel (2002) found that test-anxiety, need for approval, self-esteem, alienation, and achievement motivation all showed low relationships with cheating. More recently, the Five-Factor model of personality has been studied in relation to cheating (Nathanson et al., 2006). The Five-Factor model of personality contains the following dimensions: 1) extraversion, 2) agreeableness, 3) conscientiousness, 4) emotional stability, and 5) openness to experience. Extraversion refers to individuals who are outgoing and sociable. Agreeableness refers to individuals who are cooperative and are viewed as likeable by others. Conscientiousness refers to individuals who are responsible. Emotional stability refers to individuals who are secure and have the capacity to handle stress well. Openness to experience refers to individuals who are independent and enjoy experiencing new ventures.

Nathanson et al. (2006) states that extraversion and emotional stability have received the most attention in

relation to cheating. Specifically, in a theoretical review by Cizek (1999), in three out of four studies, extraversion showed a significant positive correlation with cheating. Yet, in a more recent study conducted by Jackson, Levine, Furnham, and Burr (2002), results showed a negative relationship between extraversion and cheating. These studies demonstrate inconsistent results in the relationship between extraversion and cheating.

The literature also reports that emotional stability has shown a very weak negative relationship with cheating (Cizek, 1999; Jackson et al., 2002). Additionally, individuals low in conscientiousness were predicted to be more likely to cheat, and in an occupational setting, these individuals tend to be more likely to engage in dishonest behaviors such as faux workers compensations claims and higher rates of absenteeism (Hogan & Hogan, 1989). Although a theoretical relationship between conscientiousness and cheating has been proposed, it has not been readily studied in the literature. Finally, agreeableness and openness to experience have not been studied in regards to their relationship with cheating nor have any theoretical connections been proposed (Nathanson et al., 2006). In a more recent study conducted by Nathanson et al. (2006), the

relationship between the big five personality dimensions and scholastic cheating were specifically addressed. Results of the first study showed that none of the Big Five personality constructs statistically predicted scholastic cheating (Nathanson et al., 2006). Results of the second study, which controlled for scholastic aptitude, showed that while extraversion, agreeableness, openness to experience, and emotional stability did not predict scholastic cheating, conscientiousness showed marginal significance with a small effect size. Nathanson et al. (2006) proposed that conscientious individuals are better prepared for examinations and are less likely to need to cheat. However, conscientious individuals are also highly ambitious, which may play a role in their desire to engage in cheating. The authors proposed that these factors imbedded within the conscientious dimensions have a "canceling effect" on one another, which may have resulted in the low significance (Nathanson et al., 2006, p. 115).

Based upon the poor predictive nature of the Big Five personality constructs, they will not be directly analyzed within this study. However, Williams et al. (2010) has provided recommendations to review the recently proposed heXaco model of personality in relation to cheating. The

heXaco Model of Personality includes an additional sixth dimension: honesty-humility, as well as the original constructs from the five factor model of personality: emotionality, eXtraversion, agreeableness, contentiousness, and openness to experience (Lee & Ashton, 2004). The HEXACO Model of Personality will be discussed in greater detail as follows.

The HEXACO Model of Personality

The goal of establishing a model of personality is to include a comprehensive framework, and the Five-Factor Model of personality is lacking in its ability to do so (Ashton & Lee, 2007, Paunonen & Jackson, 2000; Veselka, Schermer, & Vernon, 2012). In order to establish a comprehensive model of personality, "it must be based on representative samples of the universe of those characteristics" (Ashton & Lee, 2001, p. 328). Ashton an Lee (2001) state that in order to acquire such samples, personality traits must be identified by single terms in natural languages, which is known as the Lexical Hypothesis. To support this notion, lexical research seeking to identify a comprehensive personality structure has been conducted in various languages, and the major dimensions across languages have been replicated (Lee et

al., 2001). Specifically, seven languages: "Dutch, French, German, Hungarian, Italian, Korean, and Polish" have identified a six-factor structure of personality, including those originally included in the Five-Factor Model (i.e., emotional stability, extraversion, agreeableness, conscientiousness, and openness to experience), as well as an additional dimension: honesty-humility (Ashton, Goldberg, & Lee, 2004, p. 708). Based upon these findings, Ashton et al. (2001) state that the six-factor model of personality provides "the most parsimonious and comprehensive descriptive taxonomy of human personality at the subordinate level" (p. 332).

Ashton et al. (2004) state that it is quite puzzling that a six-factor structure of personality did not emerge in the English language, especially considering that the English language belongs to the Germanic branch and is influenced by French and Italian. Therefore, in attempt to uncover the inconsistency in the number of personality dimensions, Ashton et al. (2004) conducted a factor analysis based on Goldberg's (1982) set of 1,710 English personality-descriptive adjectives. Results showed that the English personality lexicon does include a sixth factor, reflecting the honesty-humility dimension found across

other languages. Ashton et al. (2004) explain that the sixth dimension may not have emerged in the English language in previous factor analyses, based upon the small number of adjectives relating to the dimension. The 1,710 adjectives provided by Goldberg (1982) more closely resembles the population of English personality-descriptors and allows for the sixth-dimension to emerge within a factor structure (Ashton et al., 2004).

The literature has provided support for a six-factor personality structure in the English language as well as across other languages throughout the world. This newly identified factor is coined in the literature as the "honesty-humility" dimension. Honesty-humility includes characteristics such as: "sincerity, fairness, modesty versus slyness, pretentiousness, and greed" (Lee et al., 2005, p. 1573). These traits closely resemble socially maladaptive characteristics- those that may play a role in cheating behaviors. Specifically, in a study by Marcus, Lee, and Ashton (2007), results indicated that the honesty dimension demonstrated strong negative correlations with self-report scholastic cheating. Additionally, Marcus et al. (2007) found that the honesty dimension also showed strong negative correlations with the dark triad

personality constructs. Williams et al. (2010) state that it would be worthwhile for future research to determine the predictive nature of the honesty-humility dimension and the "dark triad" personality constructs on cheating.

The Dark Triad

The dark triad is an emerging concept in the literature, and consists of the following personality constructs: 1) narcissism, 2) Machiavellianism, and 3) psychopathy. Characteristics of narcissism include: grandiosity, entitlement, and feelings of superiority over other individuals (Raskin & Terry, 1988). Often times individuals scoring high on narcissism scales are egocentric and arrogant (Morf & Rhodewalt, 2001). Interestingly enough, if individuals high on narcissism desire attention for their academic achievement, even if those achievements are not up-to par. Therefore, this factor may drive individuals high on narcissism to engage in cheating behaviors (Williams, Nathanson, & Paulhus, 2010).

Characteristics of Machiavellianism include: a highly pessimistic attitude and the desire to manipulate others (Christie & Geis, 1970). Individuals high on Machiavellianism often seek deceitful methods to obtain

their goals. Therefore, these factors are likely to encourage individuals with Machiavellianism tendencies to engage in cheating behaviors as well (Williams et al., 2010).

Finally, characteristics of psychopathology include four key tendencies: an unpredictable lifestyle, antisocial behaviors, manipulation, and callousness (Williams et al., 2010). Psychopathy is strongly correlated with drug and alcohol abuse, bullying, criminal acts, and behavioral cheating (Williams et al., 2004). Therefore, it is extremely likely that individuals high on psychopathy will engage in cheating behaviors on assessments. It must be noted that the use of psychopathy in this paper is *not* indicative of clinical levels.

In a study conducted by Nathanson et al. (2006), the predictive natures of the dark triad measures (i.e., narcissism, Machiavellianism, and psychopathy) were examined in relation to multiple choice cheating (copying answers from classmates). Results showed that psychopathy was the strongest predictor, and narcissism and Machiavellianism were not far behind. It must be noted that when the overlap between the three variables were controlled for in this study, psychopathy was the only

independent predictor of multiple choice cheating. Additionally, Williams et al. (2010) analyzed the association and predictive nature of the dark triad in relation to self-report cheating. Results showed that all three of the dark triad measures were positively associated with self-report cheating. Further analysis through a multiple regression (in order to show the unique effects of each member of the dark triad), indicated that psychopathy was the sole predictor of self-report cheating (Williams et al., 2010). In a second study conducted by Williams et al. (2010), the measures of the dark triad were examined in relation to behavioral cheating. Specifically, to examine behavioral cheating, student essays were submitted for plagiarism on turnitin.com (plagiarism was defined as any non-zero percentage in the study). Similar to study one, all three measures of the dark triad were significantly correlated with behavioral cheating.

The above three studies demonstrate the association between the dark triad measures and various types of scholastic cheating (i.e., multiple choice copying, self-report cheating, and essay plagiarism). Additionally, two of the above studies show the predictive nature of psychopathy in relation to scholastic cheating.

The Present Study

The present study explored the effects of cheating in UIT. Specifically, participants were randomly assigned to one of the following four conditions: 1) Low-Stakes/Instructions, 2) Low-Stakes/Non-Instruction, 3) High-Stakes/Instruction, or 4) High-Stakes/Non-Instruction (Figure 1). Cheating was detected by computing the z-statistic exemplified in the Guo et al. (2010) study in each of the four conditions for all participants. Personality characteristics (i.e., honesty, narcissism, Machiavellianism, and psychopathy) were also assessed to identify their possible link with cheating. Additionally, supplemental measures (i.e., Balanced Inventory of Desirable Responding and Mathematics Learning Questionnaire) were also assessed to identify possible links with cheating. A manipulation check for the stakes conditions was utilized to assess whether participants felt pressured to respond correctly on the CAT.

Hypotheses

Hypothesis 1

There will be a main effect between high-stakes and low-stakes conditions on cheating. Specifically, higher

levels of cheating will be found in the high-stakes condition than in the low-stakes condition.

Hypothesis 2

There will be a main effect between instruction and non-instruction conditions on cheating. Specifically, higher levels of cheating will be found in the non-instruction condition than in the instruction condition.

Hypothesis 3

The effect of high versus low-stakes conditions on cheating will vary as a function of instruction. Specifically, the presence of instructions in the high-stakes/instruction condition will result in lower levels of cheating than in the high stakes/non-instruction condition.

Hypothesis 4

There will be a significant negative relationship between honesty and cheating.

Hypothesis 5a

There will be a significant positive relationship between narcissism and cheating.

Hypothesis 5b

There will be a significant positive relationship between Machiavellianism and cheating.

Hypothesis 5c

There will be a significant positive relationship between psychopathy and cheating.

Hypothesis 6a

There will be a main effect for stakes conditions on cheating after controlling for honesty, narcissism, Machiavellianism, and psychopathology.

Hypothesis 6b

There will be a main effect for instruction conditions on cheating after controlling for honesty, narcissism, Machiavellianism, and psychopathology.

CHAPTER TWO

METHOD

Participants and Recruitment

All participants were required to be English-speaking adults over the age of 18. Participants for this study were selected through the use of convenience sampling and were recruited through a large west coast university. Upon completion of the study, participants were asked to refer others who may be interested in participating. All participants were "treated in accordance with the Ethical Principles and Code of Conduct" (American Psychological Association, 1992).

In order to determine an appropriate sample size for this study, a power analysis was conducted using G*Power Analysis (Faul & Erdfelder, 2009). Power was set at .80 and α was set at .05 in order to reach a medium effect size = .25. The results of the power analysis suggested 129 participants as an adequate sample size. The final sample included a total of 147 participants. The overall sample consisted of 77% women and 23% men with an average age of 23 years old (Table 1). The majority, 48% of students identified as Hispanic, 22% as Caucasian, 8% as African

American, 8% as Asian/Pacific Islander, 4% as Multiracial, 3% declined to state, and 1% as other (Table 1). One participant was removed as a multivariate outlier for hypotheses testing as assessed by Mahalanobis Distance (30.32) (n=146).

Measures

The measures used to collect participant information and test hypotheses included: a mock job application, the Global Cognitive Index: Quantitative Ability (developed and sold by SHL), the HEXACO-60 adapted from Ashton and Lee (2009), the Narcissistic Personality Inventory adapted from Raskin and Terry (1988), the Mach-IV Scale adapted from Christie and Geis (1970), the Levenson Self-Report Psychopathy Scale adapted from Levenson (1995), the Balanced Inventory of Desirable Responding adapted from Paulhus (1988), and the Mathematics Learning Questionnaire adapted from Fogarty and Taylor (1997).

Mock Job Application

Participants completed a mock job application as if they were applying for a job within a local organization. The items on the application included a basic demographic questionnaire, prior job experience, and educational

experience. One question administered on the verification test assessed whether or not participants felt pressured to perform well on their unproctored internet test (to measure participants' perception of stakes conditions) (Appendix C).

Global Cognitive Index: Quantitative Ability

The GCI-Q is a cognitive ability proprietary test developed and sold by SHL:

> SHL's Global Cognitive Index (GCI) with PreVisor ConVerge is a suite of cognitive ability assessments appropriate for candidates at all job levels. The Quantitative Ability assessment measures the ability to comprehend numerical information in a variety of formats. It provides an indication of how an applicant will perform when working with numbers, money, tables, bar charts, pie charts, records, analysis reports, and other data found in the workplace. The assessment provides information regarding an applicant's ability to solve math problems using basic arithmetic skills to complex algebra skills, comprehend graphs, tables, and charts, make inferences from numerical data, compare and contrast numerical data, and evaluate

quantities to arrive at a correct judgment. (SHL, 2011, p. 1).

The GCI-Q showed construct validity with the Wonderlic Classic Cognitive Ability Test (WCCAT), which is a test utilized for employee selection (r=.63). The GCI-Q also showed a meaningful correlation (r=.46) with the Global Cognitive Index-Verbal (GCI-V). The GCI-Q utilizes a stopping rule that ends the test when participants reach a standard error of .45, which is parallel to when a conventional static test reaches an internal consistency alpha of .80, indicating good reliability (Appendix D).

HEXACO-60

The HEXACO-60 was developed by Ashton and Lee (2009) and is a brief inventory that assesses six personality dimensions, including honesty-humility (H), emotionality (E), eXtraversion (X), agreeableness (A), conscientiousness (C), and openness to experience (O). The HEXACO-60 includes 10 items for each subscale (60 items in total). Participants responded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A sample item from the HEXACO-60 is "I would never accept a bribe, even if it were very large." The HEXACO-60 showed the following internal consistencies in a college sample:

honesty (α = .79), emotionality (α = .78), eXtraversion (α = .80), agreeableness (α = .77), conscientiousness (α = .78), and openness to experience (α = .77). In this specific study, the HEXACO-60 measures showed the following internal consistencies: honesty (α = .70), emotionality (α = .74), eXtraversion (α = .79), agreeableness (α = .71), conscientiousness (α = .67), and openness to experience (α = .75) (Appendix E).

The Narcissistic Personality Inventory (NPI)

The NPI was developed by Raskin and Terry (1988). The NPI is a 40-item forced choice instrument, which measures clinical criteria in the DSM-III for narcissistic personality disorder. However, this measure was designed to measure narcissism in the general population. Participants responded to forced choice items and a sample item from the NPI is: A "I insist upon getting the respect that is due to me" or B "I usually get the respect I deserve." This scale has shown good internal consistency in previous studies (α = .83). In this specific study, the NPI showed good internal consistency (α = .81)(Appendix F).

The Mach-IV Scale

The Mach-IV scale was developed by Christie and Geis (1970). The Mach-IV scale is a 20-item instrument which

measures extreme self-interest, pragmatism, and lack of trust. Participants responded to a 7-point Likert scale ranging from SD= strongly disagree to SA= strongly agree. Christie and Geis suggest that responses that range from 1-3 are considered low Mach, responses that range from 5-7 are considered high Mach, and scores of 4 are considered neutral. Christie and Gesi also suggest adding a constant of 20 to the final score, making 100 a neutral Mach score. The lowest possible Mach score is a 40, whereas the highest possible Mach score is a 160. A sample item from the Mach IV Scale is "The best way to handle people is to tell them what they want to hear." The literature supports internal consistencies of the Mach IV Scale between .70 and .80, as well as test-retest reliability (α = .76) over a six-week period (Christie, 1970; Zook, 1985; 1986). In this specific study, the Mach-IV showed adequate internal consistency (α = .76) (Appendix G).

The Levenson Self-Report Psychopathy Scale (LSRPS)

The LSRPS was developed by Levenson (1995). The LSRPS is a 26-item instrument, which measures psychopathy (i.e., a lack of empathy for others). Participants responded on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The LSRPS is comprised of both

positively and negatively worded items. A sample item from the LSRPS is "I enjoy manipulating other people's feelings." The LSRPS shows acceptable internal consistencies (α = .84 scale 1 and α = .68 scale 2) and high test-retest reliability across 8-weeks (α = .83) (Lynam, Whiteside, & Jones, 1999). In this specific study, the LSRPS showed good internal consistency (α =.84) (Appendix H).

Balanced Inventory of Desirable Responding

The BIDR was developed by Paulhus (1988). The BIDR is a 40-item instrument, which measures self-deceptive enhancement/positivity (SDE) and impression management (IM). Participants responded on a 7-point Likert scale ranging from 1 (not true) to 7 (very true). The items are comprised of both positively and negatively worded items. All 40 items may be summed to calculate an overall desirable responding measure. An example item from the SDE subscale is "I am fully in control of my own fate." A sample item from the IM subscale is "I never read sexy books or magazines." The BIDR shows good internal consistency (α = .83), and test-retest correlations (r = .69) for SDP and (r = .65) for IM over a 5-week period. The BIDR also shows concurrent validity with the Marlowe-Crowne

scale (r= .71) (Paulhus, 1988), and with the Social Desirability Inventory of Jacobson, Kellogg, Cauce, and Slavin (r= .80) (1977). In this specific study the SDP showed adequate internal consistency (α = .72) and IM showed good internal consistency (α =.84). The BIDR as a whole showed good internal consistency reliability in this sample (α =.86) (Appendix I).

Mathematics Learning Questionnaire

The Mathematics Learning Questionnaire is comprised of three sections. The first section measures confidence in mathematics. The second section measures confidence when utilizing computers. The third and final section measures attitudes towards using computers in learning mathematics. The present study only assessed the first section (i.e., math confidence). The math confidence scale is an 11-item instrument, which originates from an experimental scale by Fogarty and Taylor (1997). Participants responded on a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). The math confidence scale incorporates both positively and negatively worded items. A sample item from the Math Confidence Scale is "I do not have a mathematical mind." This scale has good internal consistency (α = .89) (Fogarty, Cretchley, Harman,

Ellerton, & Konki, 2001). In this specific study, the Mathematics Learning Questionnaire showed adequate internal consistency (α = .70) (Appendix J).

Procedure

Part 1 of the study was administered in an unproctored internet testing setting. Participants were informed that they would be completing an application and a cognitive ability test as if they were applying for a job within a local organization. Participants were provided with a link to Qualtrics, upon which, they read and agreed to an informed consent (Appendix A). Participants then filled out a mock job application asking basic demographics and prior job experience items (Appendix C). Each participant was randomly assigned to a high-stakes or low-stakes condition. In order to parallel high stakes conditions, participants were informed that the top two highest scores on the cognitive ability test would receive a cash prize (Figure 2a). Participants in the low stakes conditions were informed that two cash prizes would be raffled off with a relatively low probability of winning the raffle (Figure 2b). This method to establish high versus low stakes conditions was adopted by Bloemers, Oud, and van Dam

(2012). A total of 49% (n=71) of the sample was randomly assigned to the high stakes condition and 51% (n=75) of the sample was randomly assigned to the low stakes condition (Table 1).

Participants were then randomly assigned to an instruction or non-instruction condition. The instruction condition received the following information: (1) you will be taking a second proctored examination to verify results from the unproctored examination and (2) you will be immediately disqualified if found cheating in any form (e.g., dictionary, internet, calculator, assistance from others, etc.) (Figure 3a). The non-instruction condition was instructed that calculators will not be allowed on the test (Figure 3b). A total of 47% (n=68) of respondents were randomly assigned to the non-instruction condition, while53% (n=78) of the sample was randomly assigned to the instruction condition (Table 1).

Participants were then provided with a direct link to the cognitive ability test (GCI-Q), administered through SHL. Following the assessment portion, participants were provided with a short debriefing statement and thanked for their time (Appendix K).

Part 2 of the study was administered in a proctored test setting. Participants signed up for the proctored test session via the CSUSB SONA system web site. Participants were provided with the location, date, and time of the proctored verification test, as well as a reminder email 24 hours prior to their test. Participants met at a specified location to take the proctored verification test. Note that the test proctors of the verification test were blind to participants' condition to help control for expectancy effects. Participants were provided with and signed an informed consent form (Appendix B). Participants then completed a follow-up mock job application for consistency and data matching (i.e., through student id's and email addresses). They were then administered the GCI-Q Scale, the HEXACO-60, the Narcissistic Personality Inventory, the Mach-IV Scale, the Levinson Self-Report Psychopathy Scale, the Balanced Inventory of Desirable Responding, and the Math Confidence Scale. Following the verification process, all participants were debriefed and thanked for their time (Appendix L). Winners of the monetary incentives were announced immediately through email following the completion of the study and both winners were randomly selected. All participants were debriefed that winners of

the monetary incentives were based on a raffle rather than the top two highest scores on the selection test.

CHAPTER THREE

RESULTS

Sample

Based on the nature of the 2-part data collection procedure, 250 responses were collected for part 1 and 167 responses were collected for part 2. A total of 75 participants only participated in part 1 of the study and were removed from hypotheses testing. Additionally, another 28 participants were eliminated from hypotheses testing. Specifically, of those 28 participants, 17 did not provide any identifying information (i.e., email and/or student id) for matching purposes, and 11 participants took the unproctored test (part 1) twice. Therefore, a total of 147 participants participated in part 1 and part 2 of study and provided usable responses.

Initial analyses were conducted to assess whether there were significant differences between participants who only completed part 1 (n=75) and participants who completed both part 1 and part 2 (n=147). Participants were compared in terms of gender, ethnicity, self-report grade point average (GPA), highest education level, current employment status, and highest employment status. Significant

differences were tested utilizing the z-test of column proportions and t-tests. There were no systematic differences between participants, with the exception of highest education level. There were a greater percentage of participants (14%) who had a high school diploma for those who completed both parts of study in comparison to those who did not (4%). However, there were a greater percentage of participants (64%) who reported having some college for those who only completed part 1 of the study in comparison to those who completed both parts (46%). Specific comparison counts and percentages for demographic information may be viewed on Tables 2 and 3. Additionally, comparison counts for educational and employment information may be viewed on Tables 4 and 5.

Z-Statistic and Composite Variables

A Z-statistic (Guo et al., 2010) was calculated for each participant (criterion variable). This allowed for an overall mean Z-statistic to be calculated for each of the following four groups: 1) Low-Stakes/Instructions, 2) Low-Stakes/Non-Instruction, 3) High-Stakes/Instruction, or 4) High-Stakes/Non-Instruction. According to the one-tailed ($\alpha = .05$) z-score cut-point of 1.645, there were a total of 7%

(n=10) of the sample, who were suspected of engaging in cheating behaviors.

In order to create the composite continuous variables, only participants who completed 80% of each instrument were included. A total of 144 participants in the sample completed at least 80% of each composite continuous variable. Three participants did not complete at least 80% of the Narcissistic Personality Inventory, and were excluded based upon pairwise deletion.

Manipulation Check

Stakes Conditions

A total of 93% (n=70) of participants in the low stakes condition correctly selected that they would be entered into a $50.00 raffle in the follow-up manipulation check (Table 6). A total of 88% (n=63) of participants in the high stakes condition correctly selected that they would be given a $50.00 reward for receiving one of the top two score on the math selection test in the follow-up manipulation check (Table 6). Based upon the idea that greater than 80% of the students correctly answered the manipulation check in both the low stakes and high stakes

condition, which allows for error in students responding, all in hypotheses testing.

Instruction Conditions

A total of 88% (n=60) of participants in the non-instruction condition correctly selected that they were provided with the information that they could not use a calculator but they may use a pencil and scratch piece of paper in the follow-up manipulation check (Table 6). A total of 84% (n=65) of participants in the instruction condition correctly selected that they were provided with the instructions that they could not use a calculator or conduct internet searches, that the work they submit is their own, they do not seek or accept help in answering the questions, but they may use a pencil and scratch piece of paper in the follow-up manipulation check (Table 6). Based upon the idea that greater than 80% of the students correctly answered the manipulation check in both the non-instruction and instruction conditions, which allows for error in students responding, all participants were included in hypotheses testing.

Tests of Parametric Assumptions

Prior to completing the analyses, parametric screening was conducted on the data to assess missing data, univariate outliers, multivariate outliers, normality, homogeneity of variance, homogeneity of slope, linearity, homoscedasticity, and multicolinearity.

Missing data was analyzed through a missing values analysis. All predictor variables and the criterion were missing less than 5% of data, and therefore there was no need to remove or estimate data. There were no univariate outliers found for the stakes or instruction conditions, honesty, narcissism, Machiavellianism, or psychopathy based upon values > ±3.5 criteria and were not discontinuous from the distribution. There was one multivariate outlier assessed by Mahalanobis distance (30.32) and was removed from further analyses. Honesty, Machiavellianism and psychopathy were normally distributed as analyzed through a histogram. The narcissism variable was slightly positively skewed, which we would expect in a student population, as students often display a sense of entitlement. Homogeneity of variance was not violated for stakes or instruction conditions and was assessed based upon the SD 4:1 ratio. Homogeneity of slope was not violated for any of the

personality characteristics. Honesty, Machiavellianism, narcissism and psychopathy showed non-linear distributions with cheating and heteroscedasticity as indicated through scatter-plots. Multicollinearity was assessed for personality predictor variables through the variable inflation factor (VIF) and tolerance, no values were greater than 5 or less than .10, respectively.

Test of Hypotheses

Hypotheses 1 - 3

A 2x2 between subjects factorial analysis of variance (ANOVA) was conducted to assess significant differences between stakes and instruction conditions on cheating. The omnibus test revealed that the explained variance was not significantly greater than the unexplained variance, F (3, 142) = 1.908, $p > .05$. The effect size showed a less than small effect $\eta^2 = .039$. The effect size magnitude for hypotheses 1-3 below follow Cohen's 1988 rules of thumb.

Hypothesis 1

A main effect between high-stakes and low-stakes conditions on cheating was expected. Specifically, the high stakes condition was hypothesized to result in greater levels of cheating than the low stakes condition. A 2x2

between subjects factorial ANOVA revealed there was no main effect for the stakes condition, F (1, 142) = 1.296, $p > .05$, $\eta^2 = .009$ (less than a small effect)(Cohen, 1988). High stakes showed a mean cheating score of -.078 and low stakes showed a mean cheating score of .099. Hypothesis 1 was not supported.

Hypothesis 2

A main effect between instruction and non-instruction conditions on cheating was expected. Specifically, the non-instruction condition was hypothesized to result in greater levels of cheating than the instruction condition. The 2x2 between subjects factorial ANOVA revealed that there was no main effect for the instruction condition, F (1, 142) = .487, $p > .05$, $\eta^2 = .003$ (less than a small effect) (Cohen, 1988). The instruction condition showed a mean cheating score of .065 and the non-instruction condition showed a mean cheating score of -.043. Hypothesis 2 was not supported.

Hypothesis 3

A significant interaction was expected to result in the effects of stakes condition on cheating to vary as a function of instructions. Specifically, it was hypothesized that the high stakes/instruction combination would result

in lower levels of cheating than the high stakes/non-instruction condition. The 2x2 between subjects factorial ANOVA revealed that there was a significant interaction, $F(1, 142) = 4.471$, $p < .05$, $\eta^2 = .031$ (small effect) (Cohen, 1988). However, results showed that participants in the low-stakes/instruction condition showed a mean cheating score of (-.011), the low-stakes/non-instruction condition showed a mean cheating score of (.2081), the high stakes/non-instruction condition showed a mean cheating score of (-.295), and the high stakes/instruction condition showed a mean cheating score of (.140). Hypothesis 3 was not supported based on a significant interaction in an unanticipated direction (Figure 5).

Hypotheses 4 - 5

Zero-order correlations were assessed between the personality characteristics: honesty, narcissism, Machiavellianism, and psychopathy with the outcome variable: cheating (Figure 6). The four personality characteristics were not significantly related to cheating. Additionally, a simultaneous multiple regression was conducted on the data to assess honesty, narcissism, Machiavellianism, and psychopathy as predictors of cheating. Honesty, narcissism, Machiavellianism, and

psychopathy were non-significant predictors of cheating F (4, 138) = .621, $p > .05$, $R^2 = .018$ (less than small effect)(Figure 7).

Hypothesis 4

A significant negative relationship between honesty and cheating was hypothesized. Honesty showed a non-significant negative correlation with cheating r = -.007 (less than small effect size), $p > .05$ (Figure 6). In the regression model, honesty did not differ significantly from zero (b = -.001, β = -.001). The standardized beta weight indicates a less than small effect size (Keith, 2006).

Hypothesis 5a

A significant positive relationship between narcissism and cheating was hypothesized. Narcissism showed a non-significant correlation with cheating r = .092 (less than small effect size), $p > .05$ (Figure 7). In the regression model, narcissism did not differ significantly from zero (b = .016, β = .098). The standardized beta weight indicates a small effect size (Keith, 2006).

Hypothesis 5b

A significant positive relationship between Machiavellianism and cheating was hypothesized. Machiavellianism showed a non-significant correlation with

cheating r = -.061 (less than small effect size), p > .05 (Figure 8). In the regression model, Machiavellianism did not differ significantly from zero (b= -.168, β = -.112). The standardized beta weight indicates a moderate effect size (Keith, 2006).

Hypothesis 5c

A significant positive relationship between psychopathy and cheating was hypothesized. Psychopathy showed a non-significant correlation with cheating r = .024 (less than small effect size), p > .05 (Figure 9). In the regression model, psychopathy did not differ significantly from zero (b = .171, β = .078). The standardized beta weight indicates a small effect size (Keith, 2006).

Hypothesis 6

A 2x2 between subjects factorial analysis of covariance (ANCOVA) was conducted to assess significant differences between stakes and instruction conditions on cheating after controlling for honesty, narcissism, Machiavellianism, and psychopathy. Although the main effects for stakes and instructions conditions were not supported and honesty, narcissism, Machiavelianism, and psychopathy were not predictive of cheating, but for the purposes of the hypotheses, the ANCOVA was still performed.

The omnibus test revealed that the explained variance was not significantly greater than the unexplained variance, $F(10, 132) = 1.644$, $p > .05$, $\eta^2 = .009$ (less than small effect size) (Cohen, 1988).

Hypothesis 6a

A main effect between high-stakes and low-stakes conditions on cheating was expected after controlling for personality characteristics (i.e., honesty, narcissism, Machiavellianism, and psychopathy). Results from a 2x2 between subjects ANCOVA showed a non-significant result, $F(1, 135) = 1.030$, $p > .05$, $\eta^2 = .009$ (less than small effect) (Cohen, 1988).

Hypothesis 6b

A main effect between instruction and non-instructions conditions on cheating was expected after controlling for personality characteristics (i.e., honesty, narcissism, Machiavellianism, and psychopathy). Results from the 2x2 between subjects ANCOVA showed a non-significant result, $F(1, 135) = .789$, $p > .05$, $\eta^2 = .006$ (less than small effect) (Cohen, 1988).

Supplemental Analyses

Based upon the lack of support for hypotheses 1-6, supplemental analyses were conducted on additional measures. Results from participants' composite scores from the Balanced Inventory of Desirable Responding (i.e, self-deceptive positivity and impression management) and the Mathematics Learning Questionnaire were assessed in relation to cheating.

Supplemental Test of Parametric Assumptions

Prior to completing the analyses, parametric screening was conducted on the data to assess missing data, univariate outliers, multivariate outliers, normality, linearity, homoscedasticity and multicolinearity. Missing data was analyzed through a missing values analysis. All predictor variables and the criterion were missing less than 5% of data, and therefore did not need to be removed or estimated. There were no univariate outliers found for self-deceptive positivity, impression management, or math confidence based upon values > ±3.5 criteria. There were no multivariate outliers found assessed by Mahalanobis distance. Self-deceptive positivity and impression management were normally distributed and math confidence

was slightly negatively skewed, which would be expected in a college student sample. Linearity of the variables were assessed through a scatter-plot. Self-deceptive positivity, impression management and math confidence were non-linear. Multicollinearity was assessed through the variable inflation factor (VIF) and tolerance, no values were greater than 5 or less than .10, respectively.

Supplemental Results

Zero-order correlations were assessed between self-deceptive positivity, impression management and math learning with cheating. Self-deceptive positivity showed a significant positive relationship with cheating ($r = .165$, small effect size) (Figure 12). Impression management showed a non-significant negative relationship with cheating ($r = -.034$, less than small effect) (Figure 13). Math confidence showed a non-significant positive relationship with cheating ($r= .140$, small effect size) (Figure 14).

A simultaneous multiple regression was conducted on these data to assess self-deceptive positivity, impression management and math confidence as predictors of cheating. Self-deceptive positivity, impression management and math

confidence were non-significant predictors of cheating $F(3, 143) = 2.640$, $p > .05$, $R^2 = .052$ (small effect size) (Figure 9).

CHAPTER FOUR

DISCUSSION

Overview

The foundation for conducting this study resulted from three areas to address in the literature: first, the need to conduct a study on high and low stakes conditions utilizing a computer adaptive test, second, to test the effects of an instruction condition on cheating utilizing a control group, and third, to test the relationship between personality characteristics (i.e., honesty, narcissism, Machiavelianism, and psychopathy) and the outcome variable: cheating.

The central basis for conducting the study was designed to provide insight into environmental conditions that may provoke cheating, test methods to mitigate cheating (i.e., providing instructions and the use of computer adaptive testing), and uncover personality characteristics correlated and predictive of cheating behaviors.

Review of the Results

This discussion begins with a review of the results and is followed by an examination of the findings. The discussion will provide limitations of the present study, provide directions for applied practice as well as pose questions for future research, and end with concluding remarks.

Hypothesis 1

The high stakes condition was predicted to result in higher levels of cheating than the low stakes condition. Results did not support this hypothesis.

Hypothesis 2

The non-instruction condition was predicted to result in higher levels of cheating than the instruction condition. Results did not support this hypothesis.

Hypothesis 3

The effect of stakes conditions on cheating was expected to vary as a function of instructions. This hypothesis was not supported.

Hypothesis 4

Honesty was expected to show a significant negative relationship with cheating. This hypothesis was not

supported. Results showed a non-significant negative relationship between honesty and cheating.

Hypothesis 5a

Narcissism was expected to show a significant positive relationship with cheating. This hypothesis was not supported. Results showed a non-significant positive relationship between narcissism and cheating.

Hypothesis 5b

Machiavellianism was expected to show a significant positive relationship with cheating. This hypothesis was not supported. Results showed a non-significant *negative* relationship between narcissism and cheating.

Hypothesis 5c

Psychopathy was expected to show a significant positive relationship with cheating. This hypothesis was not supported. Results showed a non-significant positive relationship between psychopathy and cheating.

Hypothesis 6a

The high stakes condition was predicted to result in higher levels of cheating than the low stakes condition after controlling for honesty, narcissism, Machiavelianism, and psychopathy. This hypothesis was not supported.

Hypothesis 6b

The non-instruction condition was predicted to result in higher levels of cheating than the instruction condition after controlling for honesty, narcissism, Machiavelianism, and psychopathy. This hypothesis was not supported.

Examination of the Findings

Hypothesis 1

A primary goal of this project was to support the idea that participants would be more likely to cheat in high stakes than low stakes testing environments. According to the results, hypothesis 1 was not supported. Participants randomly assigned to the high stakes testing condition did not significantly differ in terms of cheating in comparison to participants randomly assigned to the low stakes condition.

According to additional analyses, a total of 64% (n=45) of participants in the high stakes condition reported that they felt pressured to respond correctly during their first selection test in comparison to 71% (n=53) of participants in the low stakes condition. This demonstrates a strong limitation to the manipulation, as

participants in the high stakes condition perceived less pressure to respond correctly and this have reduced their likelihood to engage in cheating (Table 7).

Furthermore, a total of 11% (n=8) of participants in the high stakes condition reported that they utilized outside sources (e.g., internet, calculator, friend or family member) to assist them in their first selection test, in comparison to 20% (n=15) of participants in the low stakes condition. Therefore, a much lower percentage of participants reported utilizing outside sources in the high stakes conditions. If these self-reports are accurate responses, this would support the levels of cheating in the high stakes condition as compared to the low stakes condition (Table 7).

Attention must also be drawn to the fact that this study was conducted with a college student sample. Previous research has shown that cheating is less prevalent in American colleges with ethical codes of conduct (McCabe & Trevino, & Butterfield, 2001). It is often iterated to students that cheating and plagiarism are means for expulsion. The possibility of expulsion may have been too extreme of a consequence to engage in cheating behaviors for this study, especially for a $50.00 reward. Thus,

providing support for the levels of cheating in both the high and low stakes conditions.

Hypothesis 2

A second goal of this project was to support the idea that participants would be more likely to cheat when they do not receive instructions than when they do receive instructions. According to the results, hypothesis 2 was not supported. Participants randomly assigned to the instruction condition did not significantly differ in terms of cheating in comparison to participants randomly assigned to the non-instruction condition.

According to additional analyses, a total of 68% (n=52) of participants in the instruction condition felt pressured to respond correctly during their first selection test in comparison to 68% (n=42) of participants in the non-instruction condition (Table 8). These results demonstrate that neither the instruction nor non-instruction condition felt greater pressure to respond correctly on the first selection test.

However, a total of 18% (n=14) of participants in the instruction condition reported that they utilized outside sources (e.g., internet, calculator, friend or family member) to assist them in their first selection test, in

comparison to only 13% (n=9) in the non-instruction condition. Once again, if students are honestly responding to this question, then these findings are consistent with the levels of cheating in the non-instruction and instruction conditions (Table 8).

In a study conducted by Kerkvliet and Sigmund (1999), results showed that students who were provided with verbal instructions to not engage in cheating were significantly less likely to cheat than students who were only provided with written instructions. The fact that the present study utilized written instructions to help mitigate cheating rather than verbal instructions may have reduced the effectiveness of this method.

Additionally, in a study conducted by Godfrey and Waugh (1998), results showed that instructions provided to students had no effect on actual cheating behaviors. In a more recent study conducted by Marsden, Carroll, and Neill (2005), results revealed that there was no significant association between three measures of dishonesty (i.e., cheating, plagiarism, and falsification) and students being provided with instructions and consequences for cheating.

Hypothesis 3

A third goal of this project was to support the idea that the effects of stakes conditions will vary as a function of cheating. This hypothesis was not supported. It was hypothesized that participants in the high stakes conditions would have much lower cheating scores when provided with instructions verses non-instructions. Results showed the opposite effect: participants in the high stakes condition had much lower cheating scores when they were not provided with instructions.

According to Landers and Sackett (2012), there are two opposing viewpoints on discouraging cheating behaviors. The first viewpoint is based on the premise that in essence, UIT encourages cheating. However, cheating behaviors may be discouraged through the use of warnings or even technology based verification of identity (Landers & Sackett, 2012). Previous research has demonstrated that these warnings have reduced dishonest responding on personality-based tests (Landers, Sackett, & Tuzinski, 2011). Although, these findings would suggest that warnings would also mitigate against cheating behaviors on cognitive ability or knowledge based tests, this study as well as null results from Marsden et al. (2005) show otherwise.

The second viewpoint is based on the idea that while UIT allows for cheating behaviors, discouragement of cheating may not be effective. Therefore, the primary focus for UIT researchers and practitioners should be on detecting cheating behaviors, rather than focusing efforts on discouraging cheaters. This study showed that providing instructions did not in fact discourage cheating, especially in the high-stakes condition. These results suggest that possibly alternative modes of communication about cheating need to be addressed, or simply the focus needs to shift from efforts placed on mitigating cheating to identifying cheaters. The use of the z-statistic in the Guo and Drasgow (2010) study was an effective and efficient way to detect cheating, which provides support for the use of this statistic to detect cheaters in applied settings.

Hypotheses 4 - 5

A fourth goal of this project was to support the association between personality characteristics, including: honesty, narcissism, Machiavellianism, and psychopathy with cheating. According to the results, hypotheses 4 and 5 were not supported. None of the tested personality characteristics showed a significant association with cheating.

A total of 7% (n=10) of the present sample were classified as engaging in cheating behaviors according to the z-statistic. Such a low number of cheaters in this study demonstrate truncation of range. We cannot accurately test the differences in personality characteristics (i.e., honesty, narcissism, Machiavelianism, and psychopathy) between cheaters and non-cheaters when the split is 7% to 93%, respectively.

Additionally, studying the associative and predictive nature of personality characteristics on cheating as a continuum demonstrates inherent problems. Participants on the lowest end of the cheating spectrum are likely to have experienced practice effects, followed by non-cheaters in the middle of the spectrum, and possible cheaters on the highest end of the spectrum. Those who showed practice effects and those who did not cheat (but did not show practice effects) are unlikely to show a linear trend in personality characteristics (i.e., a negative trend for honesty and a positive trend for narcissism, Machiavelianism, and psychopathy).

Hypothesis 6a - 6b

The fifth and final goal of this project was to support the idea that there would be higher levels of

cheating in the high stakes vs. low stakes condition and higher levels of cheating in the non-instruction vs. instruction condition after controlling for honesty, narcissism, Machiavellianism, and psychopathy. According to the results, hypotheses 6a - 6b were not supported.

Not only is it likely that a college student sample is unwilling to risk cheating and potentially being expelled from the University for the opportunity to win $50.00, but students completing a psychology-based research study for extra credit are likely to lack the motivation, effort, and conscientious responding as compared to a job applicant. Therefore, regardless of the incentive being offered, research studies do not often count for a grade and thus college students become less interested in the outcome. Therefore, this lends to support as to why there are such small differences when comparing both high-stakes and low-stakes environments as well as instruction and non-instruction conditions.

Additionally, the sub-dimensions of conscientiousness, such as perseverance, control, perfectionism, and industriousness may provide support for the lack of association and predictability of conscientiousness on

cheating. Specifically, the HEXACO-60 measured simple conscientiousness, rather than its sub-dimensions and may not have displayed enough sensitivity to fully capture the association with cheating, especially with a larger sample of students engaging in cheating behaviors. Therefore, a more fine-tuned analysis of conscientiousness and cheating may in fact be beneficial, especially considering mixed results from the Nathanson et al. (2006) study on conscientiousness.

Additionally, it was expected that hypotheses 6a - 6b were not supported as there were no significant differences between the high- and low-stakes conditions, instruction and non-instruction conditions, and the personality characteristics were not correlated nor predictive of cheating. The fact that very few hypotheses were supported was driven by a variety of limitations, which will be discussed as follows.

Limitations

The findings from this research should be interpreted with caution as an experimental manipulation was utilized to artificially create high vs. low-stakes conditions. The weakness of the manipulation resulted in a truncation of

range in the number of cheaters. In a true high stakes selection test setting, job candidates may be more inclined to engage in cheating behaviors. The job candidates may anticipate higher test scores and hope to enhance their probability of securing a job interview. However, it is important to note that the opposite may result, as job candidates may fear getting caught and result in undesirable consequences (Bloemers et al., 2012).

It is important to note that the sample was predominately female (77%) as compared to men (33%). Past studies have demonstrated that women are less likely to engage in cheating behaviors as opposed to men (Tibbetts, 1999). Specifically, in a past study conducted by Tibbetts (1999), men show significantly higher cheating intent and do so in part by general pleasure from engaging in such risky behaviors. Conversely, this study demonstrated that men felt less shame engaging in cheating behaviors as opposed to their female counterparts and exhibited lower self-control (Tibbetts, 1999). Therefore, a central limitation of this study may stem from the fact that there were so few male participants, which may also account for the low levels of cheating.

Additionally, this research utilized a college sample to study cheating behaviors in UIT utilizing a mock job application and selection test. It is unlikely that students responded to the selection test with the same degree of motivation and conscientious responding that a job candidate would have. Therefore, we would expect low levels of cheating, as the opportunity to win a $50.00 reward does not parallel potential employment.

Although there are distinct disadvantages of utilizing a college student sample, there are also limitations to studying cheating in UIT administered in organizations. Specifically, many organizations do not require verification testing as this adds an additional step in the hiring process and is likely to increase costs. If organizations do conduct verification proctored testing, this is often at the request of the department and only for candidates who scored high enough to move on to the next stage in the hiring process. Therefore, this makes studying cheating in UIT settings in organizations difficult to research.

Finally, the fact that only 7% of the sample engaged in cheating behaviors provided restriction of range in researching the personality characteristics of cheaters

versus non-cheaters. However, this may also be viewed as a positive limitation to organizations, as the use of a timed computer adaptive test resulted in low levels of cheating.

Future Directions for Applied Practice

There are direct benefits for organizations implementing UIT as a means for employee selection (Fetzer & Grelle, 2010). First, is the reduction of costs associated with large proctored tests sessions. Second, internet recruitment is becoming a standard by which employers must sift through applicants, as hundreds to thousands of job applicants may apply for one opening. UIT minimizes the manual labor associated with this process. Third, UIT allows organizations to attract exceptional candidates who may not actively be searching for a job, in addition to possibly enhancing the diversity pool of applicants. These benefits, for many organizations, often outweigh the potential costs associated with the administration of UIT, specifically cheating. UIT does not ensure that cheating will not occur. However, previous research has demonstrated that computer adaptive testing often helps to minimize the opportunity for test-takers to engage in cheating behaviors. This is based on the premise

that test-takers responses can be categorized into three areas: 1) correct responses from true ability of the test-taker, 2) correct responses from lucky guesses, and 3) incorrect responding resulting from carelessness and/or distraction (Liao, Ho, & Yen, 2012). The 3-parameter logistic model CAT utilized in this study corrects for lucky guesses, which demonstrating a robust method to estimate ability in comparison to classical test theory (Liao et al., 2012). Additionally, a 4-parameter model has been developed to control for careless and distracted responding. Therefore, it will be important for organizations to be mindful of which parameter model is being implemented.

Landers and Sackett (2012) also report a variety of aspects to consider when deciding to implement UIT for employee selection. The organization must determine if the use of UIT will increase the size of the applicant pool. An increase in the applicant pool is likely to result in substantial increases in the criterion, and the likelihood of hiring an appropriate fit for the position is increased. Therefore, if an organization is deciding whether or not the potential benefits of implementing UIT outweigh the consequences associated with cheating, the organization

should consider how much their applicant pool will increase.

Not only should organizations focus on whether or not they will implement an UIT, but decide if they will utilize a formula to identify potential cheaters. Specifically, in this study as well as a study conducted by Guo. et al. (2010), the z-test showed support for identifying cheaters. However, an important question for organizations must ask is what will become of candidates who are classified as cheaters as determined by a specified formula? Will cheaters be automatically disqualified from the hiring process? Will the candidates be allowed to compete in the next hurdle of the selection process? Are cheating detection formulas legally defensible in court if a potential "cheater" sues the organization for wrongful disqualification? These are all questions organizations must consider when deciding to implement an UIT and utilizing a cheating detection formula for candidate disqualification.

Future Directions for Research

In the present study a total of 7% (n=10) of the sample were classified as engaging in cheating behaviors according to the calculated z-statistic. Of those participants only 30% (n=3) admitted to utilizing outside sources such as the internet, a calculator, and/or receiving assistance from a friend or family member (Table 9). Therefore, 70% (n=7) of the participants suspected of engaging in cheating behaviors truly did not cheat or these participants did not respond truthfully. Interestingly, of the 93% of the sample who were not classified as engaging in cheating behaviors according to the calculated z-statistic, 15% (n=20) admitted to utilizing outside sources (Table 9). Therefore, these 20 participants may have randomly responded to this self-report item, or they truly engaged in cheating, yet utilizing outside sources did not assist in increasing their ability level enough to be detected. Thus, bringing this discussion to the question of: Does cheating matter?

Previous research conducted on UIT consistently concluded that about 8% of the samples engaged in cheating behaviors based upon score change evaluation (Arthur et al., 2010; Nye et al., 2008). This result is consistent

with the 7% of participants classified as engaging in cheating behaviors in the present study. However, it must be noted that score change evaluation may demonstrate an underestimation of cheating when participants cheating behaviors are not effective (Bloemers et al., 2012). Specifically, if participants' scores do not change to a certain degree, cheating behaviors may go undetected (Bloemers et al., 2012). As pointed out by Bartram (2005), it may be important to also address how cheating affects participants' scores.

In a recent study conducted by Bloemers et al., (2012), the researchers compared cognitive ability test performance between a control group and a group instructed to engage in cheating behaviors. A total of 70% of the participants in the group instructed to cheat stated that cheating was difficult and only 58% of cheaters stated they believed their cheating efforts were effective. When separated into effective and non-effective cheaters, results showed that effective cheaters performed higher than their non-effective counterparts as well as the control group.

This brings us to the question of what makes cheating effective? Does the use of multiple cheating strategies

(e.g., calculator, Google, and a friend's help) enhance effectiveness? Does motivation for the outcome (e.g., a high test score for potential employment) enhance effectiveness? Does ability in the specific subject (e.g. mathematics) enhance effectiveness but also decrease motivation to engage in cheating? In the Bloemers et al. (2012) study, the individuals who stated they were effective in their cheating behaviors performed higher than those who stated they were not effective in their cheating behaviors. How did these participants know that their cheating behaviors were effective? It may be possible that those with high self-awareness of their ability levels are consistently correlated with performance. These are all topics that may be addressed by future research. Additionally, it may be worthwhile for future research to study the effects of students' major in relation to cheating as particular majors may exhibit greater levels of pressure and result in cheating behaviors. Thus, it will also be important for future research to determine on a Likert-type scale how pressured the students felt and use this variable as a continuous predictor of cheating.

Concluding Remarks

Consistent with past studies conducted on cheating in UIT conditions, a relatively small percentage (7%) of test-takers were suspected of engaging in cheating behaviors in this study. This may be due in part by the administration of a timed computer adaptive test. Computer adaptive testing may in fact be the most reliable method to mitigate cheating to date. It may be useful for organizations to consider if the potential benefits (e.g., substantial increase in the applicant pool) outweigh the costs (i.e., cheating). Additionally, it will be important for organizations to decide if a cheating detection formula will be utilized and how potential cheaters will be handled. Finally, future researchers may want to focus on the effectiveness of various cheating strategies and self-awareness of cheating effectiveness. It is important to note that the issues and outcomes associated with cheating in UIT are largely unknown, and future research on this topic will surely benefit both research and applied practice.

APPENDIX A

INFORMED CONSENT: PART 1

Informed Consent: Part 1

You are being invited to participate in a study designed to investigate selection testing. This study is being conducted by Emily Shindledecker under the supervision of Dr. Kenneth Shultz, Professor of Psychology at California State University, San Bernardino. This study has been approved by the Department of Psychology Institutional Review Board Sub-Committee of the California State University, San Bernardino. An official Psychology IRB stamp of approval should appear on this consent form.

In part 1 of the study, you will be asked to complete a mock job application blank and a test utilized for selection purposes. The entire process of should take about 45 minutes to complete. Once you complete part 1 of the study, you will be asked to sign-up for part 2 of the study, which will take place at California State University San Bernardino. Your participation in this study is completely voluntary. You are free to withdraw from this study at any time without penalty.

Your responses from this study will be confidential. Your name will not be linked to your responses. All data will be held in the strictest of confidence in a password protected computer and by a third party testing company (SHL). There are no foreseeable risks or discomforts to participants other than those that may be experienced in daily life. Direct benefits to participants include receiving 4 units of research credit for your participation (for participating in both parts of the research study), to be used in a Psychology course of your choice (at the instructor's discretion), and an opportunity to win a monetary reward.

The results of this study will be reported in group format only. You may receive the results of this study upon completion. Results may be found at California State University, San Bernardino in SBS 541. If you have any questions or concerns regarding this research, please feel free to contact Dr. Kenneth Shultz at (909) 537-5484 or at kshultz@csusb.edu. You may also contact the Human Subjects office at California State University, San Bernardino if you have any questions or concerns about this study.

By continuing on in this study, you acknowledge that you have read the informed consent form. By clicking on the "I Agree" button below, you understand the nature and purpose of this research, and freely consent to participate.

I acknowledge that I have been informed of, and understand the

true nature and purpose of this study, and I freely consent to participate. I acknowledge that I am at least 18 years of age. Please indicate your desire to participate by placing and "X" on the line below.

Participant's X _____

Date: _____

California State University Psychology Institutional Review Board Sub-Committee			
Approved	1/11/13	Void After	1/11/14
IBB #	H-12FA-24	Chair	

APPENDIX B

INFORMED CONSENT: PART 2

Informed Consent: Part 2

You are being invited to participate in a study designed to investigate selection testing. This study is being conducted by Emily Shindledecker under the supervision of Dr. Kenneth Shultz, Professor of Psychology at California State University, San Bernardino. This study has been approved by the Department of Psychology Institutional Review Board Sub-Committee of the California State University, San Bernardino. An official Psychology IRB stamp of approval should appear on this consent form.

In part 2 of the study, you will be asked to complete a proctored mock job application blank, a test utilized for selection purposes, and personality self-report measures. The entire process of should take about 1 hour to complete. Your participation in this study is completely voluntary. You are free to withdraw from this study at any time without penalty.

Your responses from this study will be confidential. Your name will not be linked to your responses. All data will be held in the strictest of confidence in a password protected computer and by a third party testing company (SHL). There are no foreseeable risks or discomforts to participants other than those that may be experienced in daily life. Direct benefits to participants include receiving 4 units of research credit for your participation (for participating in both parts of the research study), to be used in a Psychology course of your choice (at the instructor's discretion) and the opportunity to win a monetary reward.

The results of this study will be reported in group format only. You may receive the results of this study upon completion. Results may be found at California State University, San Bernardino in SBS 541. If you have any questions or concerns regarding this research, please feel free to contact Dr. Kenneth Shultz at (909) 537-5484 or at kshultz@csusb.edu. You may also contact the Human Subjects office at California State University, San Bernardino (909) 537-7588 if you have any questions or concerns about this study.

By continuing on in this study, you acknowledge that you have read the informed consent form. By clicking on the "I Agree" button below, you understand the nature and purpose of this research, and freely consent to participate.

I acknowledge that I have been informed of, and understand the true nature and purpose of this study, and I freely consent to participate. I acknowledge that I am at least 18 years of age.

Please indicate your desire to participate by placing and "X" on the line below.

Participant's X _____

Date: _____

California State University Psychology Institutional Review Board Sub-Committee			
Approved	1/11/13	Void After	1/11/14
IBB #	H-12FA-24	Chair	

APPENDIX C

MOCK JOB APPLICATION

1. Please enter your student identification number below:

2. Please enter your email address below (so we may contact you if you are selected for an interview):

3. Gender

    a. Male

    b. Female

4. Age

5. Ethnicity

    a. Asian/Pacific Islander

    b. African American

    c. Caucasian/White

    d. Hispanic

    e. Multiracial

    f. Other

    g. Decline to state

6. Highest Education

    a. Some High School

    b. High School Diploma

c. Some College

d. Associate's Degree

e. Bachelor's Degree

f. Master's Degree

g. Doctoral Degree

7. Estimated GPA (if applicable)

8. How much work experience (in years and months) do you have?

9. Are you currently employed?

10. If you have been previously employed, indicate the level at which you work (or did work) within the organization:

a. Entry Level

b. Middle Management

c. Top Management

11. What is the minimum hourly wage you would accept for this position?

12. What times are you available to work?

a. Morning

b. Afternoon

c. Nights

d. Weekends

e. Holidays

f. All of the above

13. Did you feel pressured to respond correctly? (Will only be asked on the proctored verification test).

a. Yes

b. No

14. Did you engage in cheating through any of the following ways: (1) using outside sources such as the internet or a calculator, or (2) having a friend, family member, or anyone else to help you on your examination? (Will only be asked on the proctored verification test).

APPENDIX D

GLOBAL COGNITIVE INDEX: QUANTITATIVE ABILITY

Sample 1:

Janis has an MBA and has been in school for 1/2 of her life. Carol is 30 years old and started school later, having only been in school for five years. If Janis is four years younger than Carol, how long has she been in school?

A. 13 years
B. 15 years
C. 18 years
D. 20 years
E. 26 years

Sample 2:

***All answers are indicated in bold text.**

APPENDIX E

HEXACO-60

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91,* 340-345.

Based on a Likert-Scale ranging from 1 (strongly disagree) to 5 (strongly agree).

1. I would be quite bored by a visit to an art gallery.

2. I plan ahead and organize things, to avoid scrambling at the last minute.

3. I rarely hold a grudge, even against people who have badly wronged me.

4. I feel reasonably satisfied with myself overall.

5. I would feel afraid if I had to travel in bad weather conditions.

6. I wouldn't use flattery to get a raise or promotion at work, even if I though it would succeed.

7. I'm interested in learning about the history and politics of other countries.

8. I often push myself very hard when trying to achieve a goal.

9. People sometimes tell me that I am too critical of others.

10. I rarely express my opinions in group meetings.

11. I sometimes can't help worrying about little things.

12. If I knew that I could never get caught, I would be willing to steal a million dollars.

13. I would enjoy creating a work of art, such as a novel, a song, or a painting.

14. When working on something, I don't pay much attention to small details.

15. People sometimes tell me that I'm too stubborn.

16. I prefer jobs that involve active social interaction to those that involve working alone.

17. When I suffer from a painful experience, I need someone to make me feel comfortable.

18. Having a lot of money is not especially important to me.

19. I think that paying attention to radical ideas is a waste of time.

20. I make decisions based on the feeling of the moment rather than on careful thought.

21. People think of me as someone who has a quick temper.

22. On most days, I feel cheerful and optimistic.

23. I feel like crying when I see other people crying.

24. I think that I am entitled to more respect than the average person is.

25. If I had the opportunity, I would like to attend a classical music concert.

26. When working, I sometimes have difficulties due to being disorganized.

27. My attitude toward people who have treated me badly is "forgive and forget".

28. I feel that I am an unpopular person.

29. When it comes to physical danger, I am very fearful.

30. If I want something from someone, I will laugh at that person's worst jokes.

31. I've never really enjoyed looking through an encyclopedia.

32. I do only the minimum amount of work needed to get by.

33. I tend to be lenient in judging other people.

34. In social situations, I'm usually the one who makes the first move.

35. I worry a lot less than most people do.

36. I would never accept a bribe, even if it were very large.

37. People have often told me that I have a good imagination.

38. I always try to be accurate in my work, even at the expense of time.

39. I am usually quite flexible in my opinions when people disagree with me.

40. The first thing that I always do in a new place is to make friends.

41. I can handle difficult situations without needing emotional support from anyone else.

42. I would get a lot of pleasure from owning expensive luxury goods.

43. I like people who have unconventional views.

44. I make a lot of mistakes because I don't think before I act.

45. Most people tend to get angry more quickly than I do.

46. Most people are more upbeat and dynamic than I generally am.

47. I feel strong emotions when someone close to me is going away for a long time.

48. I want people to know that I am an important person of high status.

49. I don't think of myself as the artistic or creative type.

50. People often call me a perfectionist.

51. Even when people make a lot of mistakes, I rarely say anything negative.

52. I sometimes feel that I am a worthless person.

53. Even in an emergency I wouldn't feel like panicking.

54. I wouldn't pretend to like someone just to get that person to do favors for me.

55. I find it boring to discuss philosophy.

56. I prefer to do whatever comes to mind, rather than stick to a plan.

57. When people tell me that I'm wrong, my first reaction is to argue with them.

58. When I'm in a group of people, I'm often the one who speaks on behalf of the group.

59. I remain unemotional even in situations where most people get very sentimental.

60. I'd be tempted to use counterfeit money, if I were sure I could get away with it.

APPENDIX F

THE NARCISSISTIC PERSONALITY INVENTORY

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*(5), 890-902.

Forced choice response: A or B

1A. I have a natural talent for influencing people.

1B. I am not good at influencing people.

2A. Modesty doesn't become me.

2B. I am essentially a modest person.

3A. I would do almost anything on a dare.

3B. I tend to be a fairly cautious person.

4A. When people compliment me I sometimes get embarrassed.

4B. I know that I am good because everybody keeps telling me so.

5A. The thought of ruling the world frightens the hell out of me.

5B. If I ruled the world it would be a better place.

6A. I can usually talk my way out of anything.

6B. I try to accept the consequences of my behavior.

7A. I prefer to blend in with the crowd.

7B. I like to be the center of attention.

8A. I will be a success.

8B. I am not too concerned about success.

9A. I am no better or worse than most people.

9B. I think I am a special person.

10A. I am not sure if I would make a good leader.

10B. I see myself as a good leader.

11A. I am assertive.

11B. I wish I were more assertive.

12A. I like to have authority over other people.

12B. I don't mind following orders.

13A. I find it easy to manipulate people.

13B. I don't like it when I find myself manipulating people.

14A. I insist upon getting the respect that is due me.

14B. I usually get the respect that I deserve.

15A. I don't particularly like to show off my body.

15B. I like to show off my body.

16A. I can read people like a book.

16B. People are sometimes hard to understand.

17A. If I feel competent I am willing to take responsibility for making decisions.

17B. I like to take responsibility for making decisions.

18A. I just want to be reasonably happy.

18B. I want to amount to something in the eyes of the world.

19A. My body is nothing special.

19B. I like to look at my body.

20A. I try not to be a show off.

20B. I will usually show off if I get the chance.

21A. I always know what I am doing.

21B. Sometimes I am not sure of what I am doing.

22A. I sometimes depend on people to get things done.

22B. I rarely depend on anyone else to get things done.

23A. Sometimes I tell good stories.

23B. Everybody likes to hear my stories.

24A. I expect a great deal from other people.

24B. Everybody likes to hear my stories.

25A. I will never be satisfied until I get all that I deserve.

25B. I take my satisfactions as they come.

26A. Compliments embarrass me.

26B. I like to be complimented.

27A. I have a strong will to power.

27B. Power for its own sake doesn't interest me.

28A. I don't care about new fads and fashions.

28B. I like to start new fads and fashions.

29A. I like to look at myself in the mirror.

29B. I am not particularly interested in looking at myself in the mirror.

30A. I really like to be the center of attention.

30B. It makes me uncomfortable to be the center of attention.

31A. I can live my life in any way I want to.

31B. People can't always live their lives in terms of what they want.

32A. Being an authority doesn't mean that much to me.

32B. People always seem to recognize my authority.

33A. I would prefer to be a leader.

33B. It makes little difference to me whether I am a leader or not.

34A. I am going to be a great person.

34B. I hope I am going to be successful.

35A. People sometimes believe what I tell them.

35B. I can make anybody believe anything I want them to.

36A. I am a born leader.

36B. Leadership is a quality that takes a long time to develop.

37A. I wish somebody would someday write my biography.

37B. I don't like people to pry into my life for any reason.

38A. I get upset when people don't notice how I look when I go

out in public.

38B. I don't mind blending into the crowd when I go out in public.

39A. I am more capable than other people.

39B. There is a lot that I can learn from other people.

40A. I am much like everybody else.

40B. I am an extraordinary person.

APPENDIX G

THE MACH-IV SCALE

Christie, R. and Gesi, F. L. (1970). Studies in Machiavellianism. New York: Academic Press.

Based on a Likert-Scale ranging from 1 (strongly disagree) to 7 (strongly agree).

1. Never tell anyone the real reason you did something unless it is useful to do so.

2. The best way to handle people is to tell them what they want to hear.

3. One should take action only when sure it is morally right.

4. Most people are basically good and kind.

5. It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.

6. Honesty is the best policy in all cases.

7. There is no excuse for lying to someone else.

8. Generally speaking, men won't work hard unless they're forced to do so.

9. All in all, it is better to be humble and honest than important and dishonest.

10. When you ask someone to do something for you, it is best to give the real reasons for wanting it rather than giving reasons, which carry more weight.

11. Most people who get ahead in the world lead clean, moral lives.

12. Anyone who completely trusts anyone else is asking for trouble.

13. The biggest difference between most criminals and other people is that the criminals are stupid enough to get caught.

14. Most men are brave.

15. It is wise to flatter important people.

16. It is possible to be good in all respects.

17. Barnum was wrong when he said that there's a sucker born every minute.

18. It is hard to get ahead without cutting corners here and there.

19. People suffering from incurable diseases should have the choice of being put painlessly to death.

20. Most men forget more easily the death of their father than the loss of their property.

APPENDIX H

THE LEVENSON SELF-REPORT PSYCHOPATHY SCALE

Levenson, M.; Kiehl, K.; Fitzpatrick, C. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68,* 151-158.

Based on a Likert-Scale ranging from 1 (strongly disagree) to 5 (strongly agree).

1. Success is based on survival of the fittest; I am not concerned about the losers.

2. I find myself in the same kinds of trouble, time after time.

3. For me, what's right is whatever I can get away with.

4. I am often bored.

5. In today's world, I feel justified in doing anything I can get away with to succeed.

6. I find that I am able to pursue one goal for a long time.

7. My main purpose in life is getting as many goodies as I can.

8. I don't plan anything very far in advance.

9. Making a lot of money is my most important goal.

10. I quickly lose interest in tasks I start.

11. I let others worry about higher values; my main concern is with the bottom line.

12. Most of my problems are due to the fact that other people just don't understand me.

13. People who are stupid enough to get ripped off usually deserve it.

14. Before I do anything, I carefully consider the possible consequences.

15. Looking out for myself is my top priority.

16. I have been in a lot of shouting matches with other people.

17. I tell other people what they want to hear so that they will do what I want them to do.

18. When I get frustrated, I often "let off steam" by blowing my top.

19. I would be upset if my success came at someone else's expense.

20. Love is overrated.

21. I often admire a really clever scam.

22. I make a point of trying not to hurt others in pursuit of my goals.

23. I enjoy manipulating other people's feelings.

24. I feel bad if my words or actions cause someone else to feel emotional pain.

25. Even if I were trying very hard to sell something, I wouldn't lie about it.

26. Cheating is not justified because it is unfair to others.

APPENDIX I

BALANCED INVENTORY OF DESIRABLE RESPONDING

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp.17-59). New York: Academic Press.

Based on a Likert-Scale ranging from 1 (not true) to 7 (very true).

1. My first impressions of people usually turn out to be right.

2. It would be hard for me to break out of any of my bad habits.

3. I don't care to know what other people really think of me.

4. I have not always been honest with myself.

5. I always know why I like things.

6. When my emotions are aroused, it biases my thinking.

7. Once I've made up my mind, other people can seldom change my opinion.

8. I am not a safe driver when I exceed the speed limit.

9. I am fully in control of my own fate.

10. It's hard for me to shut off a disturbing thought.

11. I never regret my decisions.

12. I sometimes lose out on things because I can't make up my mind soon enough.

13. The reason I vote is because my vote can make a difference.

14. My parents were not always fair when they punished me.

15. I am a completely rational person.

16. I rarely appreciate criticism.

17. I am very confident of my judgments.

18. I have sometimes doubted my ability as a lover.

19. It's all right if me if some people happen to dislike me.

20. I don't always know the reasons why I do the things I do.

21. I sometimes tell lies if I have to.

22. I never cover up my mistakes.

23. There have been occasions when I have take advantage of someone.

24. I never swear.

25. I sometimes try to get even rather than forgive and forget.

26. I always obey laws, even if I'm unlikely to get caught.

27. I have said something bad about a friend behind his or her back.

28. When I hear people talking privately, I avoid listening.

29. I have received too much change from a salesperson without telling him or her.

30. I always declare everything at customs.

31. When I was young I sometimes stole things.

32. I have never dropped litter on the street.

33. I sometimes drive faster than the speed limit.

34. I never read sexy books or magazines.

35. I have done things that I don't tell other people about.

36. I never take things that don't belong to me.

37. I have taken sick leave from work or school even though I wasn't really sick.

38. I have never damaged a library book or store merchandise without reporting it.

39. I have some pretty awful habits.

40. I don't gossip about other people's business.

APPENDIX J

MATHEMATICS LEARNING QUESTIONNAIRE

Fogarty, G., & Taylor, J. (1997). Learning styles among mature-age students: Some comments on the Approaches to Studying Inventory (ASI-S). *Higher Education Research and Development, 16* (3), 321-330.

Based on a Likert-Scale ranging from 1 (strongly agree) to 5 (strongly disagree).

1. I have less trouble learning mathematics than other subjects.

2. When I have difficulties with mathematics, I know I can handle them.

3. I do not have a mathematical mind.

4. It takes me longer to understand mathematics than the average person.

5. I have never felt myself able to learn mathematics.

6. I enjoy trying to solve new mathematics problems.

7. I find mathematics frightening.

8. I find many mathematics problems interesting and challenging.

9. I don't understand how some people seem to enjoy spending so much time on mathematics problems.

10. I have never been very excited about mathematics.

11. I find mathematics confusing.

APPENDIX K

DEBRIEFING: PART 1

Debriefing: Part 1

Part 1 of the study you have just completed was designed to examine unproctored internet testing.

Please log-on to your CSUSB SONA account to sign-up to participate in part 2 of the study. Part 2 of the study will consist of a proctored test taking place at California State University, San Bernardino. If you do not have a CSUSB SONA account, please email me at: shindlee@coyote.csusb.edu to set-up a date/time to sign-up to participate in part 2 of the study.

I am currently working towards completion of my Master's Thesis at California State University, San Bernardino and greatly appreciate your participation. Please do not discuss the contents of this study with others. The group results of this study will be available in September, 2013. If you have any questions, comments, or concerns about the study, or would like to obtain the group results of this study please feel free to contact Emily Shindledecker at shindlee@csusb.edu or Dr. Kenneth Shultz at (909) 537-5484 or kshultz@csusb.edu.

APPENDIX L

DEBRIEFING: PART 2

Debriefing: Part 2

Part 2 of the study you have just completed was designed to examine the factors associated with inconsistent responding in unproctored versus proctored internet testing. This study will provide organizations with insight into unproctored internet testing.

Please note that cash prizes will not be awarded to the two highest scores on the examination, as this was simply a manipulation in the study. However, all participants will be entered into a cash prize raffle and winners will be contacted by email following the completion of the study.

I am currently working towards completion of my Master's Thesis at California State University, San Bernardino and greatly appreciate your participation. Please do not discuss the contents of this study with others. The group results of this study will be available in September, 2013. If you have any questions, comments, or concerns about the study, or would like to obtain the group results of this study please feel free to contact Emily Shindledecker at shindlee@csusb.edu or Dr. Kenneth Shultz at (909) 537-5484 or kshultz@csusb.edu.

APPENDIX M

TABLES

Table 1

Random Assignment into Stakes and Instruction Conditions

Condition	Count	%
Stakes Condition		
High Stakes	71	49%
Low Stakes	75	51%
Instruction Condition		
Instruction	78	53%
Non-Instruction	68	47%

Table 2

Demographic Information for Participants who Completed Part 1 Only

Demographic Information	Count	%
Gender		
Women	56	75%
Men	19	25%
Ethnicity		
African American	3	4%
Asian/Pacific Islander	8	11%
Caucasian	16	21%
Hispanic	42	56%
Multiracial	2	3%
Other	2	3%
Decline to State	2	3%

Note: Mean age of sample = 22 years old.

Table 3

Demographic Information for Participants who Completed Both Part 1 and Part 2

Demographic Information	Count	%
Gender		
Women	113	77%
Men	34	23%
Ethnicity		
African American	14	10%
Asian/Pacific Islander	12	8%
Caucasian	32	22%
Hispanic	79	54%
Multiracial	7	5%
Other	1	1%
Decline to State	2	1%

Note: Mean age of sample = 23 years old.

Table 4

Educational and Employment Information for Participants who Completed Part 1 Only

Education & Employment	Count	%
Highest Education		
Some High School	1	1%
High School Diploma	3	4%
Some College	48	64%
Associate's Degree	17	23%
Bachelor's Degree	5	8%
Master's Degree	-	-
Doctoral Degree	1	1%
Currently Employed		
Yes	49	66%
No	25	34%
Highest Employment Level		
Entry Level	42	69%
Middle Management	19	31%
Top Management	-	-

Note: Mean GPA of sample = 3.20.

Table 5

Educational and Employment Information for Participants who Completed Both Part 1 and Part 2

Education & Employment	Count	%
Highest Education		
Some High School	2	1%
High School Diploma	20	14%
Some College	67	46%
Associate's Degree	39	27%
Bachelor's Degree	18	12%
Master's Degree	-	-
Doctoral Degree	-	-
Currently Employed		
Yes	82	57%
No	62	43%
Highest Employment Level		
Entry Level	101	77%
Middle Management	30	23%
Top Management	-	-

Note: Mean GPA of sample = 3.13.

Table 6

Manipulation Check: Correct Condition Identification

Condition	Count	%
Stakes Condition		
High Stakes	63	88%
Low Stakes	70	93%
Instruction Condition		
Instruction	65	85%
Non-Instruction	60	88%

Table 7

Stakes Conditions Self-Report on Cheating

Condition	Count	%
Pressured to Respond Correctly		
High Stakes	45	64%
Low Stakes	53	71%
Reported Engaging in Cheating Behaviors		
High Stakes	8	11%
Low Stakes	15	20%

Table 8

Instruction Conditions Self-Report on Cheating

Condition	Count	%
Pressured to Respond Correctly		
Instruction	52	68%
Non-Instruction	42	68%
Reported Engaging in Cheating Behaviors		
Instruction	14	18%
Non-Instruction	9	13%

Table 9

Self-Report on Cheating

Condition	Count	%
Reported Engaging in Cheating Behaviors		
Cheaters	3	30%
Non-Cheaters	20	15%

Table 10

Correlations Between Honesty, Narcissism, Machiavelianism, Psychopathy and Cheating

Variables	Honesty	Narcissism	Mach	Psych	Cheating
Honesty	1				
Narcissism	-0.203*	1			
Mach	-0.528*	0.088	1		
Psych	-0.584*	0.049	0.533*	1	
Cheating	-0.007	0.092	-0.061	0.024	1

Significant at the $p < .05$ value.

Table 11

Correlations between Self-Deceptive Positivity, Impression Management, Math Confidence and Cheating

Variables	SDP	IM	Math	Cheating
SDP	1			
IM	0.451*	1		
Math	0.162*	-0.015	1	
Cheating	0.165*	-0.034	0.14*	1

Significant at the $p < .05$ value.

APPENDIX N

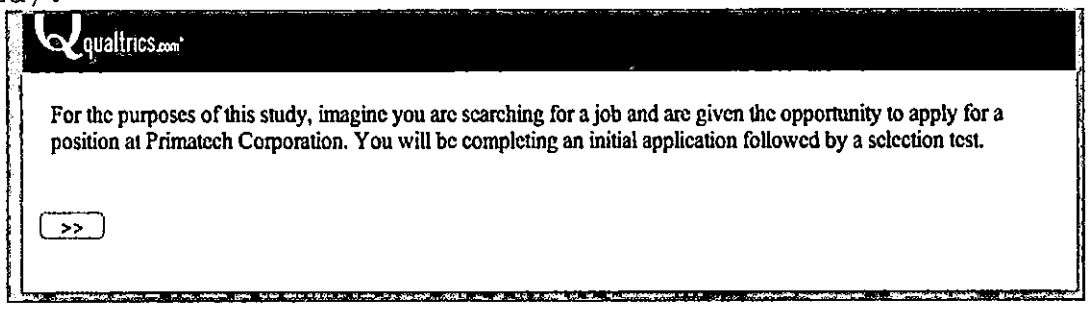
FIGURES

Figure 1. 2x2 Between Subjects Factorial ANOVA

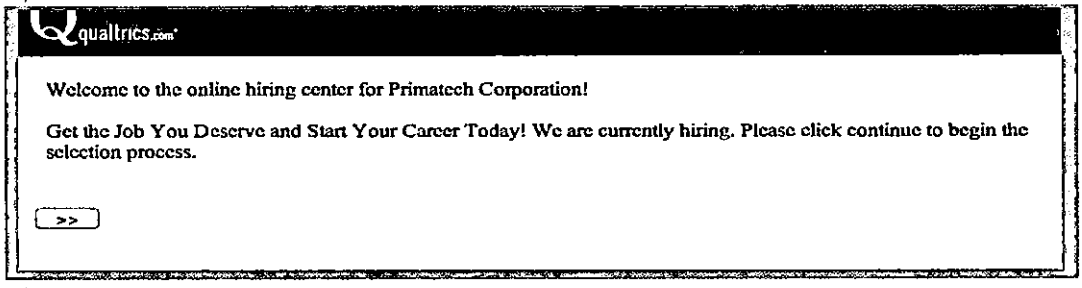
Low-Stakes/Instruction	**High-Stakes**/Instruction
Low-Stakes/**Non-Instruction**	**High-Stakes/Non-Instruction**

Figures 2a-2d. Introduction Slides

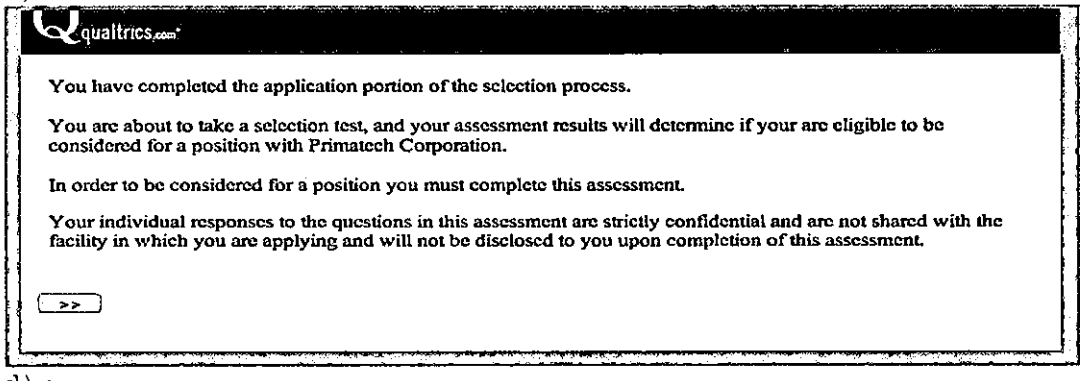
(2a):



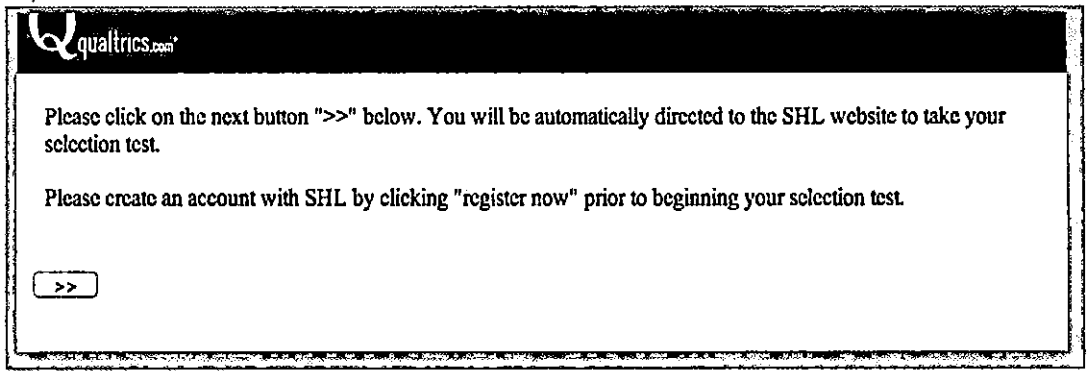
(2b):



(2c):

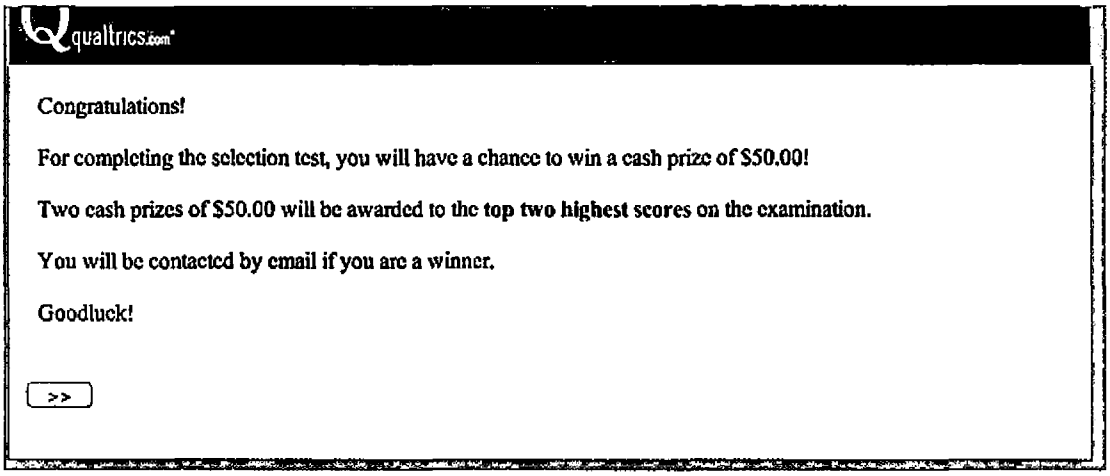


(2d):

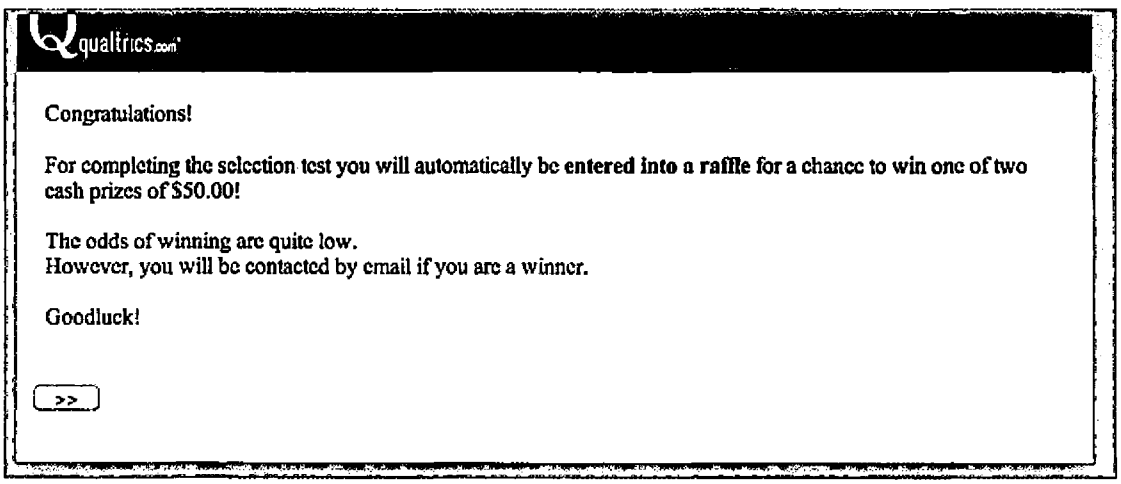


Figures 3a-3b. High Stakes versus Low Stakes Conditions

High Stakes (3a):

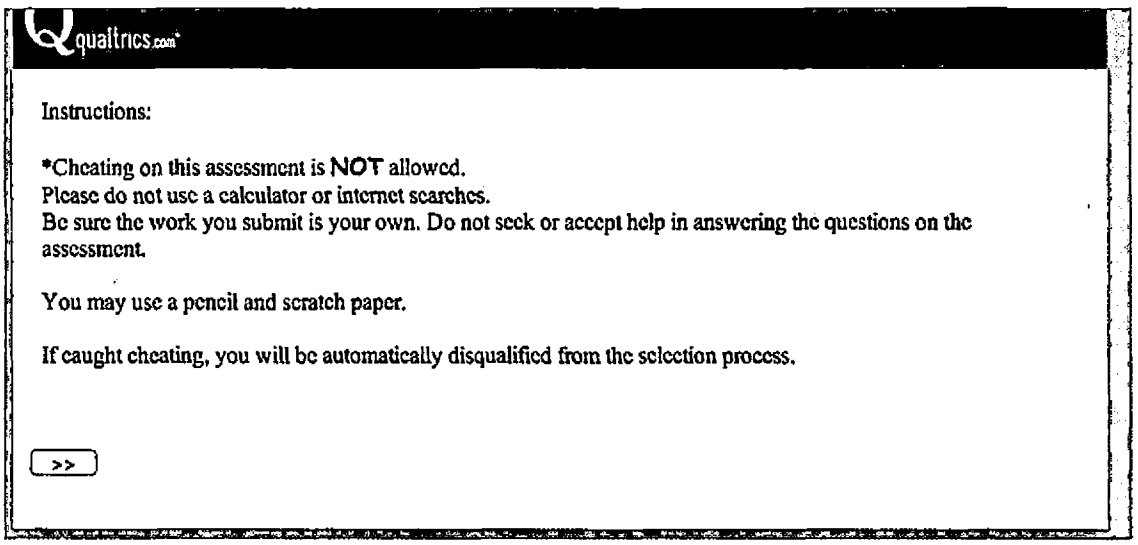


Low Stakes (3b):



Figures 4a-4b. Instruction and Non-Instruction Conditions

Instruction Condition (4a):



Non-Instruction Condition (4b):

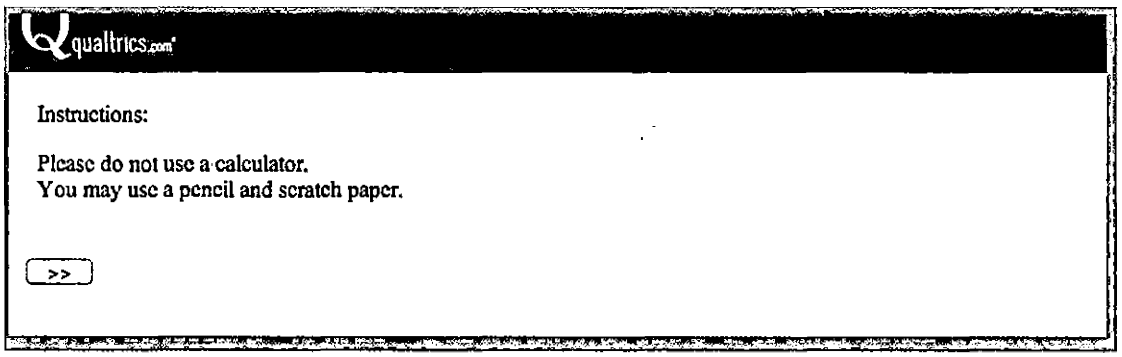


Figure 5. Effects of Stakes and Instructions on Cheating

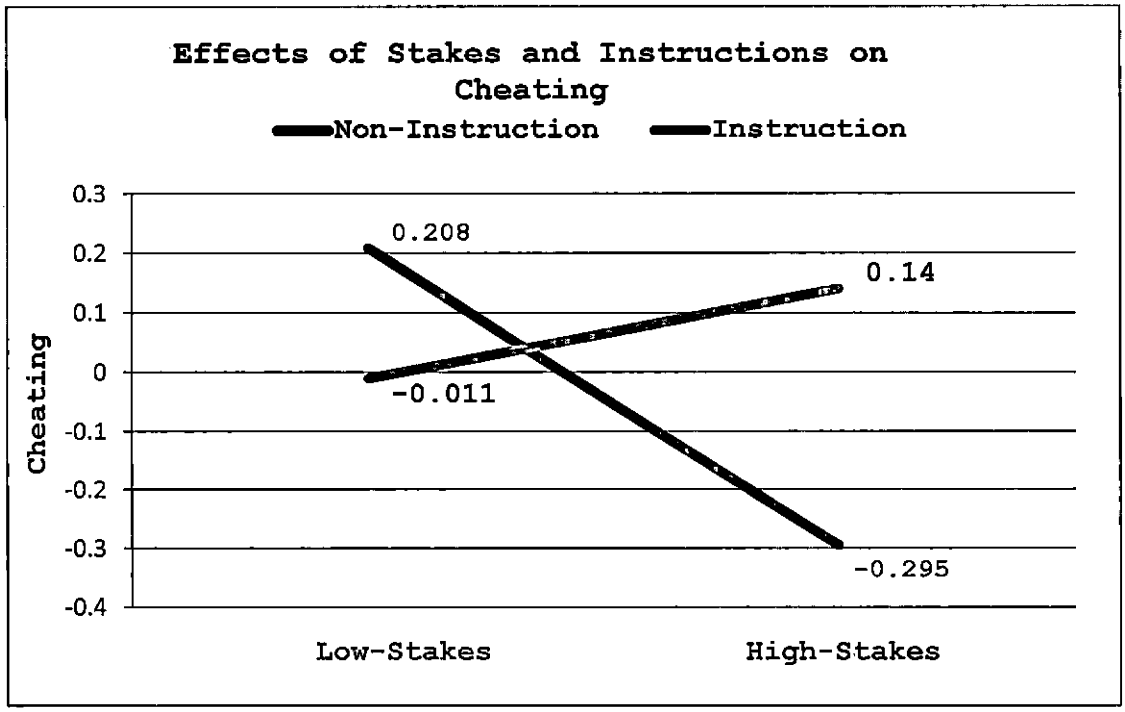


Figure 6. Correlation between Honesty and Cheating

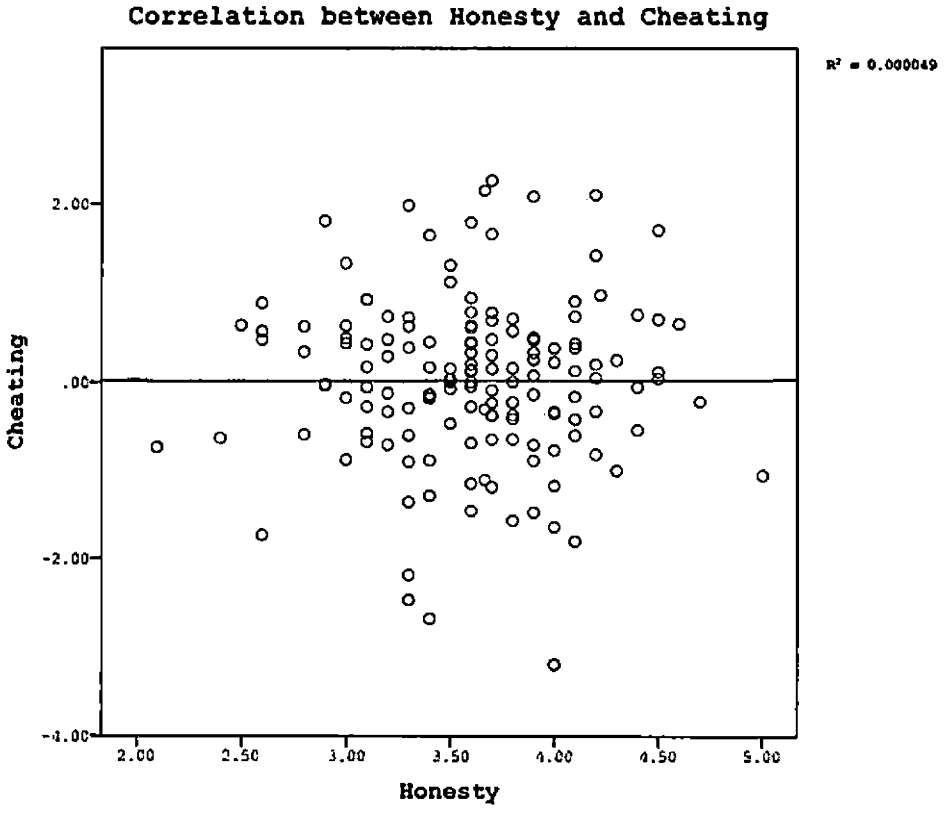


Figure 7. Correlation between Narcissism and Cheating

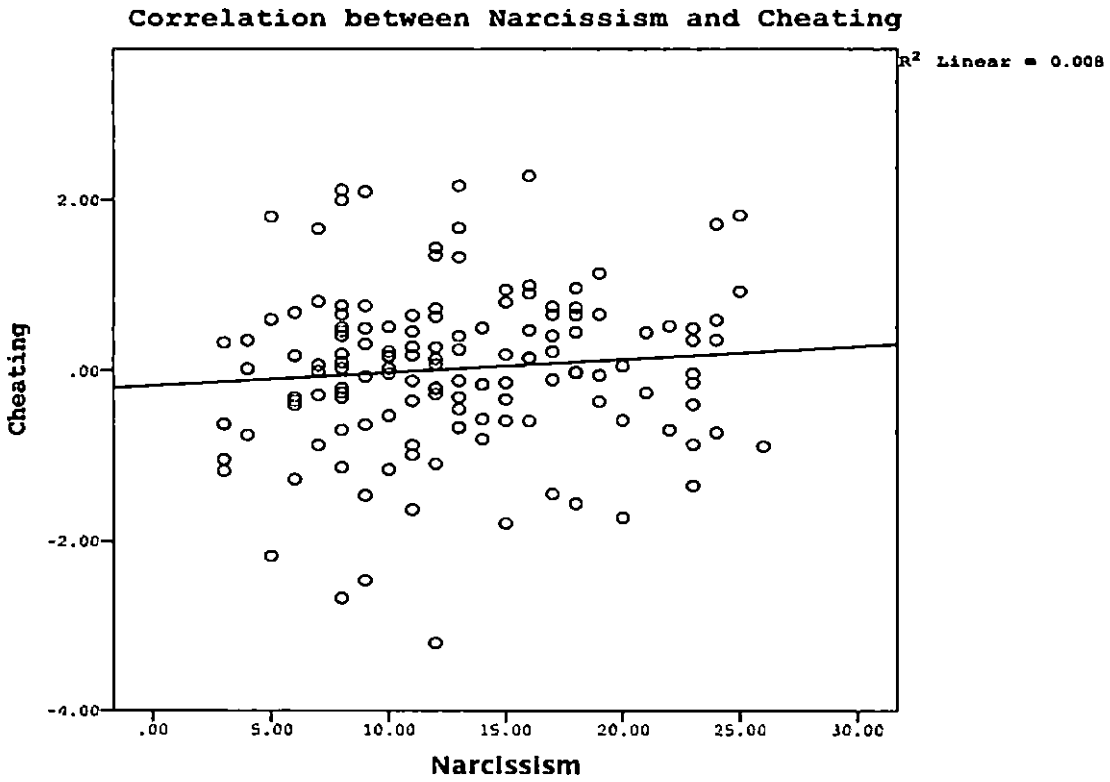


Figure 8. Correlation between Machiavelianism and Cheating

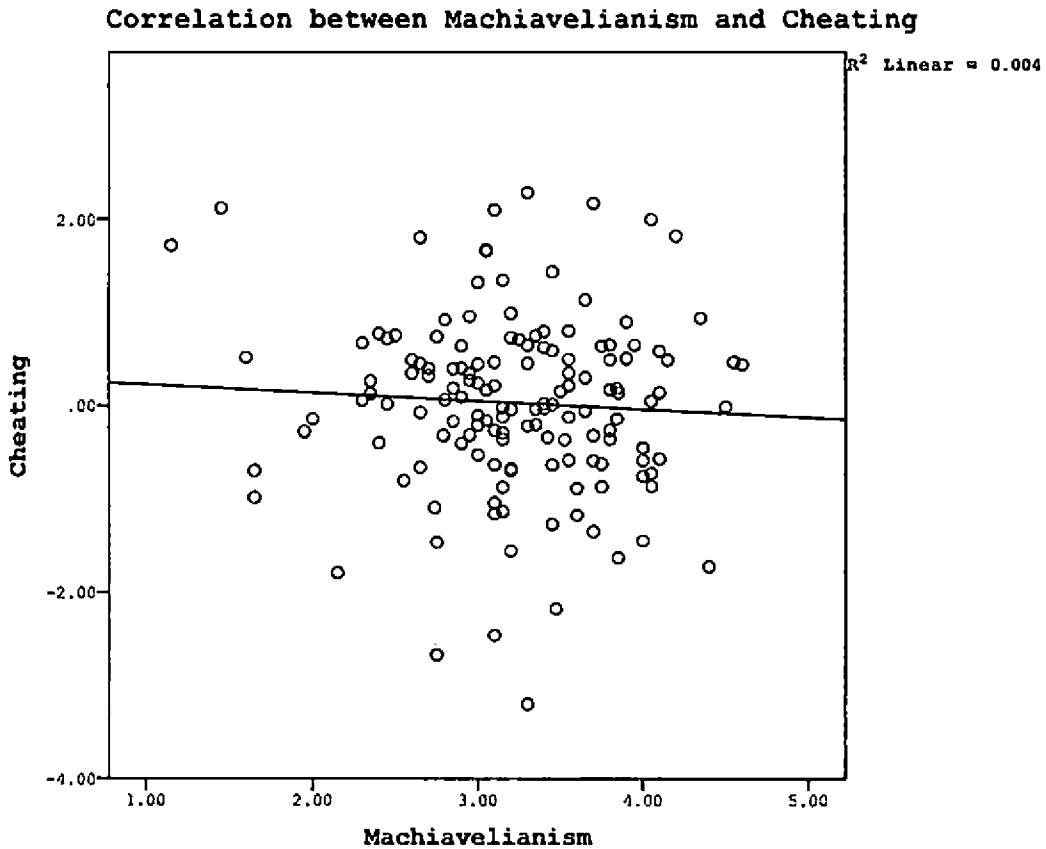


Figure 9. Correlation between Psychopathy and Cheating

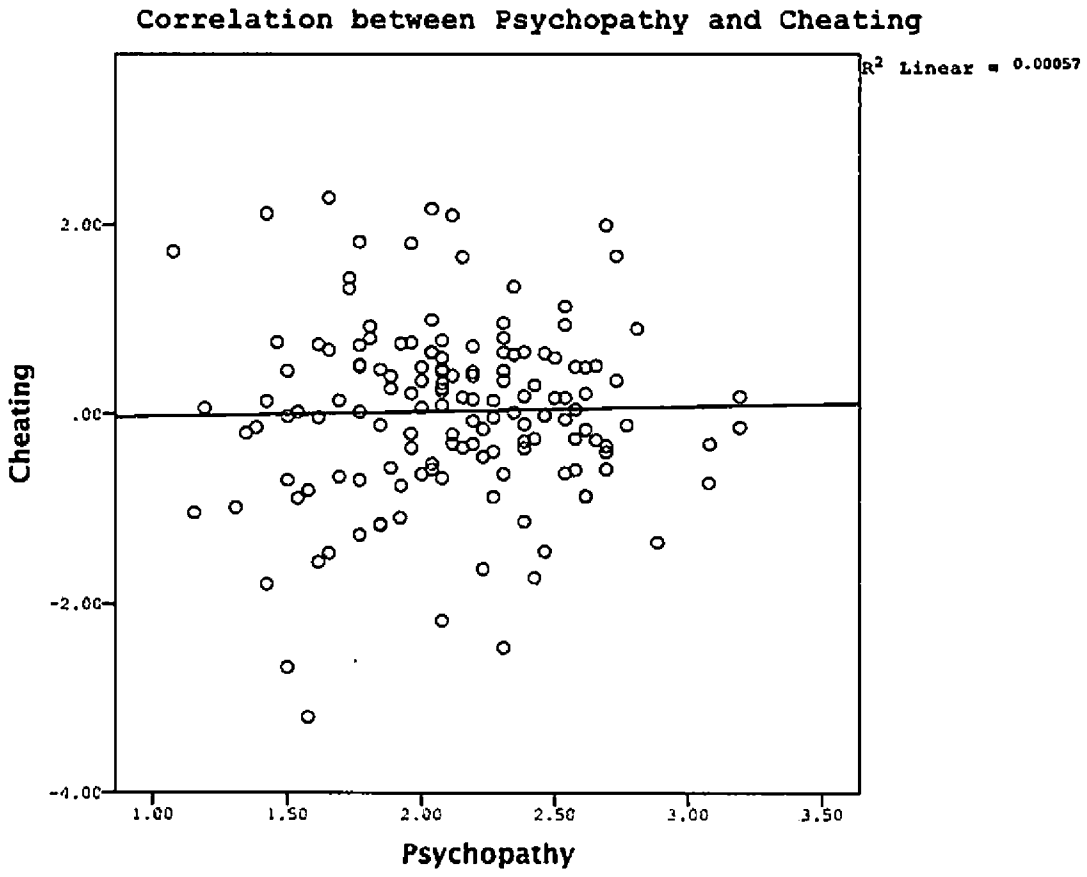


Figure 10. Honesty, Narcissism, Machiavelianism and Psychopathy as Predictors of Cheating

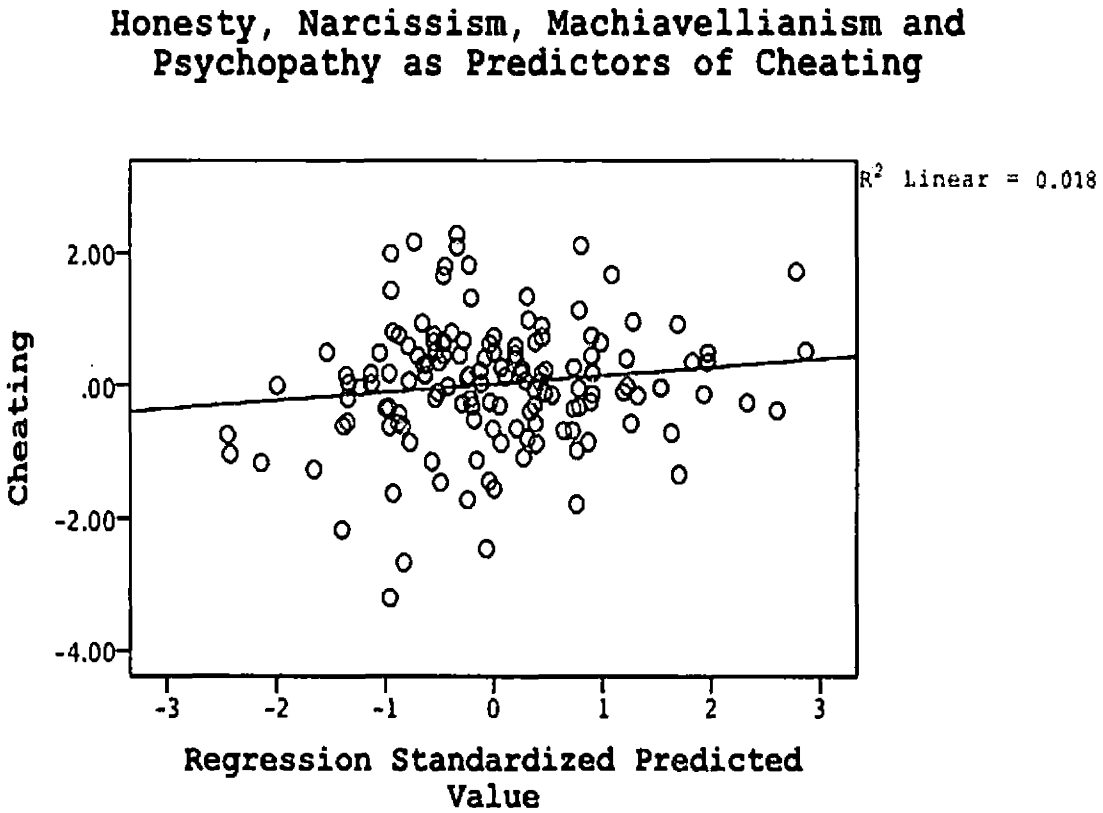


Figure 11. Correlation between Self-Deceptive Positivity and Cheating

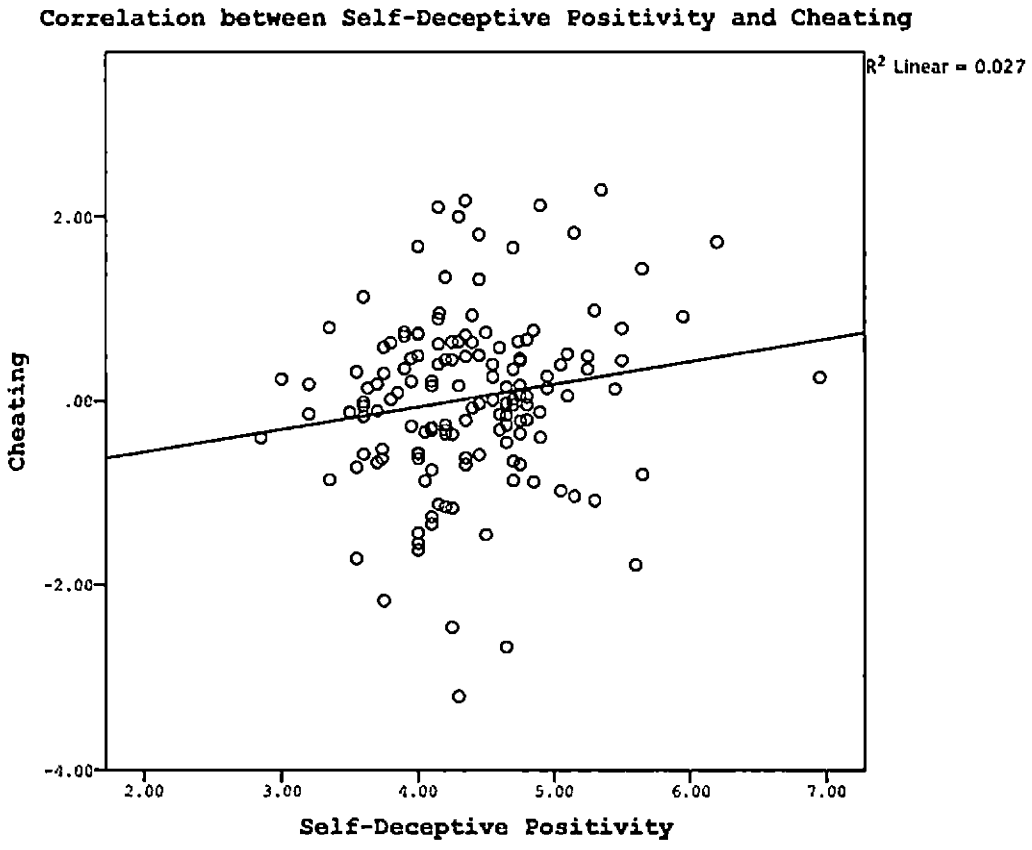


Figure 12. Correlation between Impression Management and Cheating

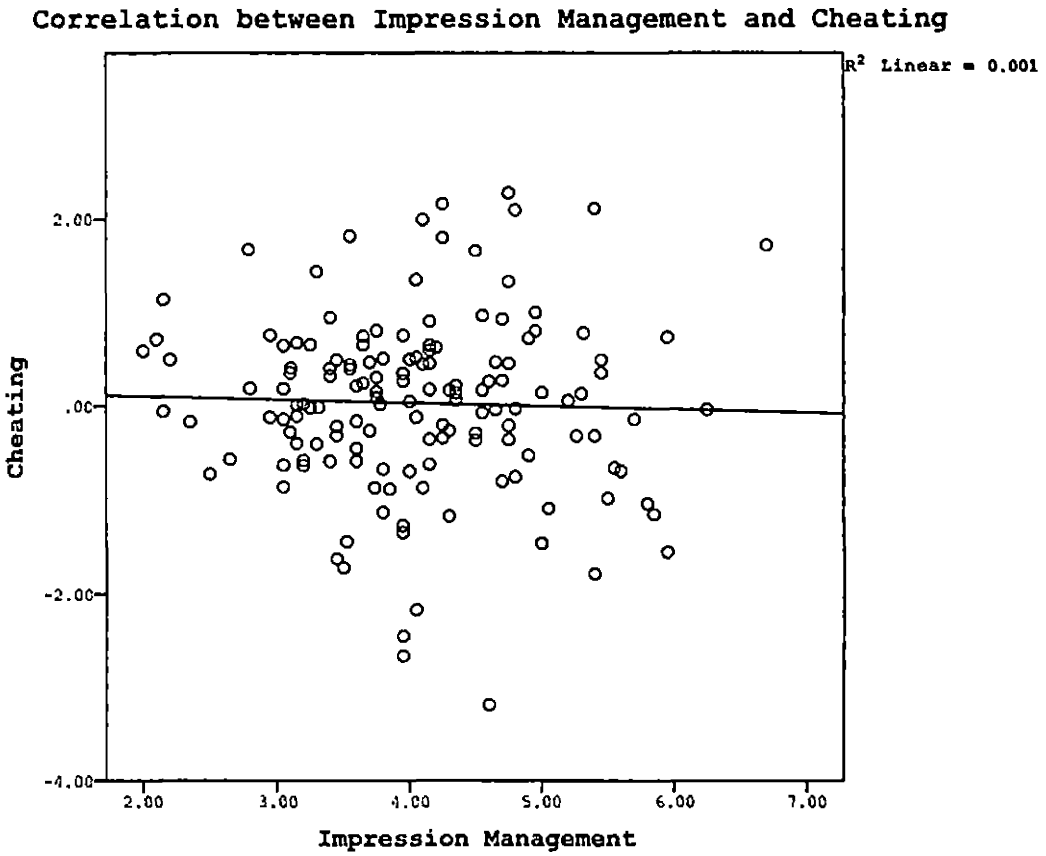


Figure 13. Correlation between Math Confidence and Cheating

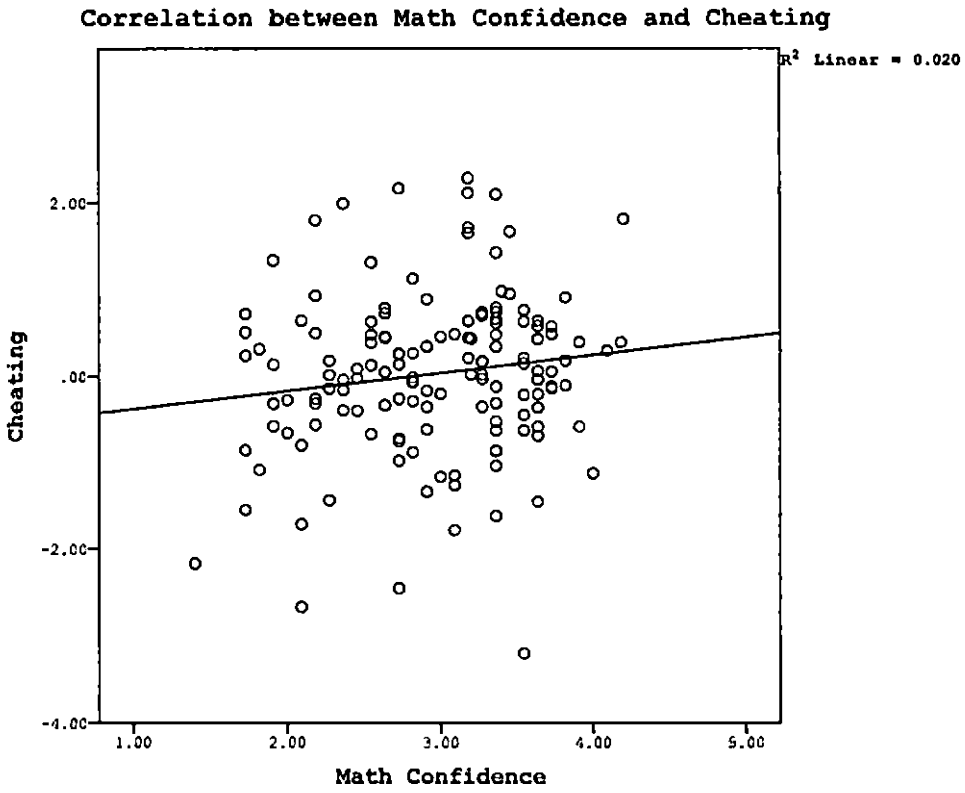
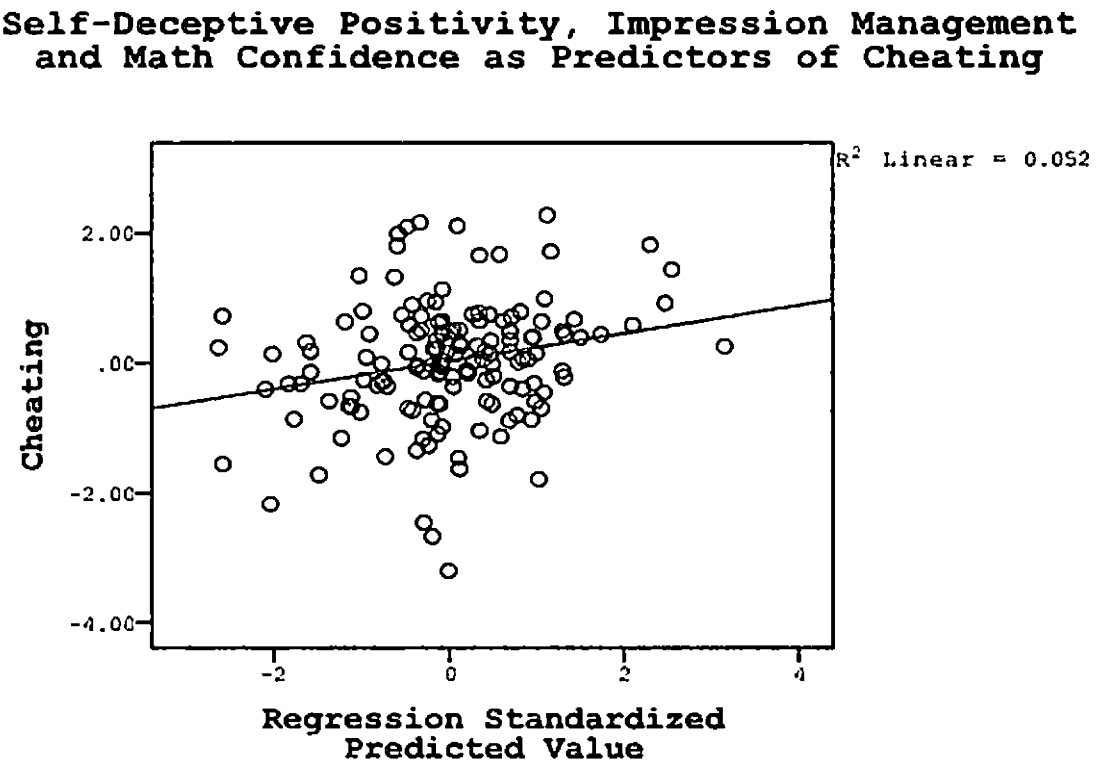


Figure 14. Self-Deceptive Positivity, Impression Management and Math Confidence as Predictors of Cheating



REFERENCES

American Psychological Association,(1992). Ethical principles of psychologists and code of conduct. *American Psychologists, 47,* 1597-1611.

Anderson, J. (2012). SAT and ACT to tighten rules after cheating scandal. *The New York Times.* Retrieved from http://www.nytimes.com/2012/03/28/education/after-cheating-scandal-sat-and-act-will-tighten-security.html? r=2&pagewanted=all

Anderson, N. (2005). *Relationships between practice and research in personnel selection: Does the left hand know what the right is doing?* In the *Blackwell Handbook of Personnel Selection,* eds. A. Evers, N. Anderson and O. Voskuijl, Malden, MA: Blackwell, pp. 1-24.

Arthur, W. R., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2009). Unproctored Internet-based tests of cognitive ability and personality: Magnitude of cheating and response distortion. *Industrial And Organizational Psychology: Perspectives On Science And Practice, 2*(1), 39-45.

Arthur, W. R., 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.

Ashton, M. C., Goldber, L. R., & Lee, K. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology, 87*(5), 707-721.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345.

Bartram, D. (2001). The Development of International Guidelines on Test Use: The International Test Commission Project. *International Journal of Testing, 1*, 33-53.

Beaty, J. C., Nye, C. D., Borneman, M. J., Kantrowitz, T. M., Drasgow, F., & Grauer, E. (2011). Proctored versus unproctored Internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal Of Selection And Assessment, 19*(1), 1-10.

Bloemers, W, Oud, A., & van Dam, K. (2012). Unproctored cognitive ability internet testing: Does cheating pay

off? *Manuscript submitted to the Journal of Applied Psychology.*

Christie, R. and Geis, F. L. (1970). Studies in Machiavellianism. New York: Academic Press. Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it.* Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Do, B. (2009). Research on unproctored Internet testing. *Industrial And Organizational Psychology: Perspectives On Science And Practice, 2*(1), 49-51.

Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial And Organizational Psychology: Perspectives On Science And Practice, 2*(1), 46-48.

Fallaw, S. S., Solomonson, A. L., & McClelland, L. (2009). *Current trends in assessment use: A multi-organizational survey.* Poster submitted for the 24th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009).

Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

Fetzer, M., Grelle, D. (2010). PrVisor ConVerge. The best Practice for unproctored/unsupervised internet testing. *SHLPreVisor White Paper 2010,* 1-8.

Fetzer, M., Dainis, A., Lambert, S., & Meade, A. (2011). Computer Adaptive testing (CAT) in an employment context. *SHLPreVisor White Paper 2011,* 1-14.

Fogarty, G., & Taylor, J. (1997). Learning styles among mature-age students: Some comments on the Approaches to Studying Inventory (ASI-S). *Higher Education Research and Development, 16* (3), 321-330.

Fogarty, G. J., Cretchley, P., Harman, C., Ellerton, N., Konki, N. (2001). Validation of a questionnaire to measure mathematics confidence, computer confidence, and attitudes to the use of technology for learning mathematics. *Mathematics Education Research Journal, 13,* 154-160.

Goldberg, L. R. (1982). From ace to zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), Advances in personality assessment (Vol. 1, pp. 203-234).

Hillsdale, NJ: Erlbaum.

Guest, D., Michie, J., Conway, N., & Sheehan, M. (2003). Human resource management and corporate performance in the UK. *British Journal of Industrial Relations, 41,* 291-314.

Guo, J., & Drasgow, F. (2010). Identifying Cheating on Unproctored Internet Tests: The Z□test and the likelihood ratio test. *International Journal Of Selection And Assessment, 18*(4), 351-364.

Guo, J., Tay, L., & Drasgow, F. (2009). Conspiracies and test compromise: An evaluation of the resistance of test systems to small-scale cheating. *International Journal of Testing, 9,* 283-309.

Harel, G., & Tzafrir, S. (2001). HRM practices in the public and private sectors: Differences and similarities. *Public Administration Quarterly, 25,* 317-350.

Hense, R., Golden, J. H., & Burnett, J. (2009). Making the case for unproctored Internet testing: Do the rewards outweigh the risks?. *Industrial And Organizational Psychology: Perspectives On Science And Practice, 2*(1), 20-23.

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology, 74,* 273-279.

Honan, W. H. (1995). Computer admissions test to be given less often. The New York Times, pp. A16.

Huselid, M.A. (1995). The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. *Academy of Management Journal, 38,* 635-672.

Jackson, C. J., Levine, S. Z., Furnham, A., & Burr, N. (2002). Predictors of cheating behavior at a university: A lesson from the psychology of work. *Journal of Applied Social Psychology, 32,* 1031-1048.

Jacobsen, L. I., Kellogg, R. W., Cauce, A. M., & Slaven, R. S. (1977). A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society, 9,* 109-110.

Kantrowitz, T. M., & Dawson, C. R. (2011). Computer adaptive testing (CAT): A faster, smarter, and more secure approach to pre-employment testing. *Journal of Business Psychology, 26,* 227-232.

Keith, T. (2006). *Multiple regression and beyond.* Pearson Education Inc.

Kerkvliet, J., & Sigmund, C. L. (1999). Can we control cheating in the classroom? *Journal of Economic Education, 30,* 331-343.

Landers, R. N., & Sackett, P. R., (2012). Offsetting Performance losses due to cheating in unproctored internet-based testing by increasing the applicant pool. *International Journal of Selection and Assessment, 20*(2), 220-228.

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology, 96,* 202-210.

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences, 38,* 1571-1582.

Lee, K., & Ashton, M. C. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340-345.

Liao, W., Ho, R., & Yen, Y. (2012). The four parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses.

Social Behavior and Personality, 40(10), 1679-1694.

Lievens, F., & Burke, E. (2011). Dealing with the threats inherent in unproctored internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology, 84*, 817-824.

Levenson, M.; Kiehl, K.; Fitzpatrick, C. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology, 68*, 151-158.

Lynam, D. R., Whiteside, S., & Jones, S. (1999). Self-reported psychopathy: A validation study. *Journal of Personality Assessment, 73*(1), 110-132.

Makransky, G., & Glas, C. W. (2011). Unproctored internet test verification: Using adaptive confirmation testing. *Organizational Research Methods, 14*(4), 608-630.

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition. *Personnel Psychology, 60,* 1-34.

Marsden, H., Carroll, M., & Neill, J. T., (2005). Who cheats at university? A self-report study of dishonest

academic behaviours in a sample of Australian university students. *Australian Journal of Psychology, 57*(1), 1-10.

McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Dishonesty in academic environments: The influence of peer reporting requirements. *Journal of Higher Education, 72,* 29-45.

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry, 12,* 177-196.

Naglieri, J. A., Drasgow, F., Schmite, M., Handler, L., Prifitera, A., Margolis, A. & Velasquez, R. (2004) Psychological Testing on the Internet: New problems, old issues. *American Psychologist, 59,* 150-162.

Nathanson, C., Paulhus, D. L., Williams, K. M. (2006). Predictors of a behavioral measure of Scholastic cheating: Personality and competence but not demographics. *Contemporary Educational Psychology, 31,* 97-122.

Nye, C. D., Do, B., Drasgow, F., & Fine, S. (2008). Two-step testing in employee selection: Is score inflation a problem?. *International Journal Of Selection And Assessment, 16*(2), 112-120.

Paulhus, D. L. (1988). Assessing self deception and impression management in self-reports: the balanced inventory of desirable responding. (Manual available from the author at the Department of Psychology, University of British Columbia, Vancouver, B.C., Canada V6T 1Y7).

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp.17-59). New York: Academic Press.

Paulhus, D. L., Nathanson, C., & Williams, K. M. (2005). A new look at the link between cognitive ability and exam cheating. In B. D. Zumbo (Ed.), *Developments in the theories and applications of measurement, evaluation, and research methodologies across the disciplines*. Vancouver, BC: University of British Columbia Press.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality, 68,* 821-835.

Pearlman, K. (2009). Unproctored Internet testing: Practical, legal, and ethical concerns. *Industrial And*

Organizational Psychology: Perspectives On Science And Practice, 2(1), 14-19.

Pew Internet and American Life Project (2007) Demographics of Internet Users. Retrieved March 16, 2012, from http:// www.pewinternet.org/trends/User Demo 1.11.07.htm.

Raskin, R., & Terry, H. (1988). A principal components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54,* 890-902.

Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving Personnel Selection Practices in New Zealand Organisations and Recruitment Firms. *New Zealand Journal of Psychology, 31,* 8-18.

Terpstra, D.E., & Rozell, E.J. (1993). The Relationship of Staffing Practices to Organizational Level Measures of Performance. *Personnel Psychology, 46,* 27-48.

Tibbetts, S. G. (1999). Differences between women and men Regarding decisions to commit test cheating. *Research In Higher Education, 40*(3), 323-342.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored internet testing in employment settings.

Personnel Psychology, 59 (1), 189-225.

Veselka, L., Schermer, J. A., Vernon, P. A. (2012). The dark triad and an expanded framework of personality. *Personality and Individual Differences, 55,* 417-425.

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education,* 39, 235-274.

Whitley, B. E., Jr., & Keith-Spiegel, P. (2002). Academic dishonesty: An educator's guide. Mahwah, NJ: Erlbaum.

Williams, K. M., Nathanson, C., & Paulhus, D. L. (2010). Identifying and profiling scholastic cheaters: Their personality, cognitive ability, and motivation. *Journal of Experimental Psychology: Applied, 16*(3), 293-307.

Wolf, A., & Jenkins, A. (2006). Explaining Greater Test Use for Selection: The Role of HR Professionals in a World of Expanding Regulation. *Human Resource Management Journal, 16,* 193 - 213.

Zook, A., II, (1985). On measurement of Machiavellianism. *Psychological Reports, 57,* 982.

Zook, A., II, & Sipps, G. J. (1986) Reliability data and sex differences with a gender-free Mach IV. *Journal of Social Psychology, 126*(1), 131-132.