

Emotional Reactions to a Middle Warning Message: A Physiological Approach

by

Elissa M. Hack

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama

May 5, 2018

Keywords: middle warning, faking, dark triad,
personality, autonomic response

Approved by

Jinyan Fan, Chair, Associate Professor of Psychology

Jennifer Robinson, Associate Professor of Psychology

Daniel Svyantek, Professor of Psychology

Ana Franco-Watkins, Professor of Psychology

Abstract

Socially desirable responding or faking has posed a serious threat to the validity and utility of personality tests used as selection tools for organizations. There have been numerous attempts to circumvent this issue. A recently suggested method involves implementing a warning message during the middle of the testing process and then allowing the test-taker to re-test. There is empirical evidence that such a middle warning may lead to score reduction, resulting in more accurate scores for fakers, but less accurate scores for non-fakers. What has not been thoroughly examined, however, is the underlying emotional process behind this observed effect. The proposed study aims to investigate the emotional mechanisms through the measurement of physiological autonomic responses in conjunction with a self-report emotions survey.

Participants in this study were 244 college students enrolled at a Southeastern university in the U.S. The current study uses a 2×2 factorial design with fakers vs. non-fakers and warning vs. control as the two factors. Participants were attached to physiological recording equipment and were then instructed to complete the following personality questionnaire with either honest or fake-good instructions. A warning and control message was randomly delivered in the middle of the online personality test resulting in an initial test and re-test. This study primarily examined four emotions: fear, guilt, anxiety, and anger.

Table of Contents

Abstract.....	ii
List of Tables	vi
List of Figures.....	vii
Introduction.....	1
Previous Studies.....	3
Measuring Faking	3
Managing Faking	6
Middle Warning Studies	8
Current Study	13
Contributions.....	13
Study Design.....	14
Emotions.....	15
Physiology	20
Method	31
Participants	31

Procedure	31
Measures	36
Results.....	39
Analytic strategies.....	39
Manipulation check.....	41
Over compensatory response check.....	42
Hypothesis testing.....	46
Expected physiological response	48
Mediation	51
Moderated meditation.....	52
Qualitative results	57
Discussion.....	60
Manipulation and over compensatory check	60
Emotional reaction to warning message	62
Physiological support.....	63
Mediations.....	63
Moderated mediations.....	64
Contributions	65

Limitations and future research	66
References.....	68
Appendix A	76
Appendix B	80

List of Tables

Table 1	30
Table 2	42
Table 3	43
Table 4	44
Table 5	44
Table 6	47
Table 7.1	50
Table 7.2	50
Table 8	51
Table 9	53
Table 10	54
Table 11	54
Table 12	55
Table 13	56
Table 14	56

List of Figures

Figure 1	64
----------------	----

Introduction

Due to the ease of administration, lack of adverse impact as compared to cognitive measures, and capability of predicting various measures of performance, the use of personality inventories in personnel selection has become increasingly popular among employers over recent years (Barrick & Mount, 1991; Bobko, Roth, & Potosky, 1999; Barrick, Mount, & Judge, 2001; Hough & Oswald, 2008). In conjunction with their increasing popularity, however, there has been concern regarding the relative ease to which faking behavior can arise on these non-cognitive measures as well as how this faking behavior can impact the validity of scores in a negative manner (Ellingson, Sackett, & Hough, 1999; Viswesvaran & Ones, 1999). Faking has been defined as “the tendency to deliberately present oneself in a more positive manner than is accurate in order to meet the perceived demands of the testing situation” (Fan, Gao, Carroll, Lopez, Tian, & Meng, 2012, p.867).

In response to this there have been numerous strategies to attempt to address these issues of faking. The following literature review will outline the various strategies that research has utilized to attempt to manage dishonest responding; specifically a warning message component. I will then expand upon the literature regarding the use of a warning message delivered during the middle of the test that allows the respondents a chance for retest, as this will be the format used for the current study. There has been little research in regards to the possible mechanisms behind the middle warning component; so I will outline the emotions literature and postulate which discreet emotions I believe most likely at the moment of the warning message. Lastly I will

outline how I can use physiological measurements in conjunction with emotions surveys to advance the literature regarding warning messages as deterrents for faking.

Previous Studies

Measuring Faking

Researchers have attempted to deal with the issue of socially desirable responding (SDR) in regards to personality testing for the past 60 years (Edwards, 1957). There have been numerous scales and techniques developed to attempt to address this prevalent issue (Wiggins, 1968). Many of the first approaches to address the issue of SDR involved statistical correction, response sets analyses, and rating test items for social desirability (Crowne & Marlowe, 1960). One of the first of these approaches was the SD (social desirability) scale developed by Edwards (1957). However, as the scale was based on the clinical MMPI, many scholars began to question the content of the item pool and its theoretical background (Wiggins, 1968). In 1960, Crowne and Marlowe developed their highly utilized scale of social desirability that did not include the psychopathology items previously included in Edwards (1957) SD scale. The exclusion of items traditionally used to measure psychopathology, encouraged the use of SDR scales outside of the clinical setting and made their use more legally defensible in an organizational setting. Paulhus (1984) argued for examining socially desirable responding in terms of impression management and self-deception and how they relate to the Alpha and Gamma factors. The Alpha factor, or unconscious bias, relates to Edward's (1957) SD scale, and the Gamma factor, representing deliberate falsification, closely relates to Wiggin's (1959) Social Desirability scale (Paulhus, 1984). Since this initial exploration, there have been numerous theories and models put forth as to the underlying structure of socially desirable responding (for a full review see Paulhus, 2002).

Many of these models centered around two underlying factors split between a conscious vs. unconscious bias.

Since this time, Paulhus' Impression Management (IM) and Self-Deceptive Enhancement (SDE) measures have been commonly used in an attempt to detect faking or socially desirable responding on personality measures. Originally, the construct of SDE, unconscious bias, concerned the test-taker's tendency to characteristically view oneself in a positive light. IM, on the other hand, involved the deliberate, or conscious, effort to falsify one's responses to the test in order to gain an advantageous position (Barrick & Mount, 1996). For example, an SDE item may state, "I never regret my decisions," and the applicant would endorse the extent to which this was true. Most individuals at one time or another have regretted a decision, and therefore any endorsement stating otherwise can be considered socially desirable responding. Similarly, an example of an IM item may state, "I always tell the truth." This is most certainly not true for any person, so stating firm agreement with this statement can also be considered socially desirable responding. Paulhus has since proposed a new model of socially desirable responding that incorporates a content-level and process level to result in 4 types of SDR (Paulhus, 2002). The self-deception component is now split into enhancement and denial, while the impression management component is now divided based upon the concepts of agency and communion (Paulhus, 2002).

There, however, has been some debate as to whether these types of questions are actually indicative of dishonest responding. The format of the questions does not allow for a firm statement that an applicant was intentionally being deceitful. An applicant could truly make a point to never regret anything in their life and therefore could be punished for responding honestly. This had led to an exploration for alternative measures of applicant faking.

Two such alternative measures are the Bogus Statement inventory (Anderson, Warner, & Spencer, 1984; Dwight & Donovan, 2003) and Over-Claiming Questionnaires (OCQ; Paulhus, Harms, Bruce, & Lysy, 2003). Unlike IM and SDE scales, these measures have distinctly false and distinctly true items. Bogus Statement (BS) inventories involve a series of job related tasks that an individual can claim various levels of experience or knowledge. There are, however, non-existent tasks mixed in with genuine tasks. An example of a bogus statement, in which one would claim a level of experience with, would be something similar to, “Matrixing solvency files” (Anderson, Warner, & Spencer, 1984). This is a non-existent task; therefore, if the individual claims knowledge of this procedure, they are shown to be responding in a socially desirable manner. BS items are more targeted towards the construct of impression management, as the results are outright falsifications. Over-claiming questionnaires (OCQ) are very similar to BS items in that there is a mixture of true and non-existent elements that the individual claims familiarity with. OCQs, however, deal with general knowledge, such as people, events, and things, rather than specifically job-related items. An example of a non-existent OCQ would be to rate one’s familiarity with, “cholarine” (Paulhus, Harms, Bruce, & Lysy, 2003).

These new measures, while promising, have not yet been fully established as valid. There has also been recent support for the validity of using the traditional IM and SDE measures of SDR. Both IM and SDE have shown that they have a significant moderation effect on personality scores, supporting their relevance in regards to personality and faking (Hack, Yao, Page, Yuan, & Fan, 2015). This study found that IM scores moderated criterion-related validities of Agreeableness, Openness, and Conscientiousness-ambition scores, whereas SDE scores moderated criterion-related validities of Agreeableness, Emotional Stability, and Conscientiousness-dutifulness scores in a real-world selection context. Therefore, this study relies

on impression management as a manipulation check to measure desirable responding. Impression management is particularly susceptible to instructed faking, and is a conscious choice in response patterns on the part of the participant (Paulhus, 2002). Self-deceptive enhancement, on the other hand, is defined as an unconscious response bias and was not found to be sensitive to specific faking instructions (Paulhus, 2002). As this proposed study uses an instructed faking condition, impression management is the most ideal measure to use as a manipulation check for faking with this particular study design.

Managing Faking

There are primarily three categories in which faking mitigation procedures can be defined: reactive, proactive, and a combination of the two (Fan et al., 2012). The above four measurements of faking (BS, OCQ, IM, SDE) can be used with the reactive strategy, in which faking is allowed to occur and then controlled for after-the-fact. Statistical correction and statistical modeling can also fall within this tactic.

In contrast to the reactive strategy, there has been an increase interest in the research of attempting to proactively address faking in the form of preventative measures. Examples of this strategy include a pre-warning message, subtle items, and forced choice format (Fan et al., 2012). A pre- warning message attempts to address the issue of faking through warning test takers against dishonest responding before they begin the testing process. Subtle items attempt to prevent the applicant from identifying the specific criteria being measured and forced-choice formats attempt to avoid socially desirable responding by making all answers equally desirable (Christiansen, Burns, & Montgomery, 2005). As this study concentrated on warning messages, I will first describe the various standard pre-warning messages and then introduce the third faking mitigation strategy, as it is what I employed for my research.

The traditional format for presenting a warning to deter faking has been constructed using an identification and consequence component. Identification informs the test taker that there is an embedded social desirability measure that has the ability to detect dishonest responding, while the consequences component informs the test taker that there will be penalties if they are identified as fakers such as removal from the application process (Pace & Borman, 2006). In more recent literature, there has been an effort towards warning components that may be less aversive in regards to the test takers' perceptions of organizational justice. Pace and Borman (2006) identify three additional warning types to both detection and consequences. The "appeal to reason" approach involves encouraging individuals to respond in an honest manner because it is in both their and the companies best interest in order to find the best suited match. An "educational" warning, attempts to convey honest and open communication from the organization to encourage honest responding in return. Lastly, an "appeal to morals" warning message encourages the test takers to view themselves as trustworthy and authentic individuals in an attempt to gain honest responses from the attempt to remain consistent with this perception.

These warning messages tend to have a more friendly tone than that of the more traditional approaches (Pace & Borman, 2006). An article by Dullaghan (2010), however, found that the detection and consequences warning was more effective in deterring faking, and did not have a significant impact on perceived procedural justice. Therefore, there has been a new approach to examining warnings in which the component of the warning message is not the focus, but rather the timing (Butcher, Morfitt, Rouse, & Holden, 1997; Landers, Sackett, & Tuzinski, 2011; Burns, Fillipowski, Morris, & Shoda, 2015).

Middle Warning Studies

More recently, there has been promising evidence for administering a warning component in the middle of the testing process as opposed to the more traditional pre-warning setup. This procedure is a type of the third faking mitigation strategy, which combines the components of the reactive and proactive categories. It is reactive in that it allows faking to occur in an initial block while measuring faking and proactive by then warning the individual that faking has been detected and give them a chance for recourse. The first research to explore this avenue was conducted by Butcher, Morfitt, Rouse, and Holden (1997). These researchers explored re-testing after a warning using the MMPI-2 in a real-world selection context. The MMPI-2 is a personality test that included multiple validity scales imbedded within. This study in particular focused on the K (defensiveness), and L (lie) scales to determine those that would be retested. Test-takers were applicants for an airline pilot position and were given the MMPI-2 as part of their pre- screening evaluation. If they were found to have a T score above a cutoff score on K or L their test was considered invalid and they were given a second MMPI-2 with further instructions resembling an identification and appeal to reason warning message. Of the 72 applicants whose tests were considered invalid upon initial testing, 57 produced valid profiles after retesting. The warning component as retesting resulted in more valid and interpretable results lending preliminary support to this new method (Butcher et al., 1997).

Landers, Sackett, and Tuzinski (2011) chose a similar mid-warning setup for their field study of applicants for managerial positions at a national retailer. These researchers examined the effects of warning and retesting on what they term “blatant extreme responding (BER).” The study had an impressive sample size of 32,311 with 20,993 who completed the test after the warning was implemented. Applicants who completed the measure within the first 13 months of

testing only received a general warning message against faking before the testing process. Those who completed the measure after the 13-month mark were given the same pre-warning but were also given a pop-up warning based upon %100 BER for the first 1/3 of the test. This warning consisted of only an identification component. They were then given the opportunity to go back and change their answers accordingly. From this procedure stemmed three separate groups: those who had never answered with BER, those who had responded with BER and continued to do so, and those who initially responded with BER but changed their answers after receiving the warning message. The results showed a significant decrease in BER suggesting that a real-time warning does reduce applicants who have been responding in a socially desirable manner as well as confirming that faking does occur and matter in the real world selection context. There is an issue, however, with using BER as an indication of socially desirable responding in that there is no direct evidence that faking has occurred.

Another study conducted by Fan et al. (2012) also found that introducing a warning component after an initial testing block and giving individuals a chance for recourse caused a significant decrease in faker's personality scores upon retesting. The first study sample consisted of 157 real-world applicants for staff positions at a university in China. These applicants were given a personality assessment, which included a bogus statement and impression management scale, as well as a measure of test fairness, face validity, test satisfaction, and test motivation. If the applicants were found to exceed the faking criteria after the initial block of the test, they were given a warning message and given the chance to retest. Those who were not flagged received a control message. The results showed that flagged applicants reduced their scores significantly, while the non-flagged applicants score did not vary significantly from time 1 to time 2. To perform a true experiment, the researchers conducted a second study, which randomly assigned

the warning and control message to applicants. The results demonstrated that those who were given the control message did not greatly reduce their scores regardless of their having been flagged. Those who received the warning message reduced their scores but the reduction was significantly greater for those who had been flagged as fakers. These results also lend support to the notion of a mid-test warning message to reduce the likelihood of faking. One issue that was not examined, however, was the mechanisms behind why test takers will reduce their scores after receiving this warning message. Could it be due to guilt from having lied, or fear that a continuation in their previous response pattern would remove them from the selection process?

In a very similar study to Fan et al. (2012), Ellingson, Heggstad, and Makarius (2012) examined whether retesting did lead to a more accurate representation of flagged individuals' true scores, as well as examined what emotional reactions were occurring at the moment of retest. Unlike the previous studies, this research was conducted using undergraduate students in the lab setting. There was, however, a deception component to attempt to motivate socially desirable responding from those who would do so in a real application setting. In the first study participants completed a baseline personality measure and trait guilt measure before the deception component. They were then given the deception and told to do their best on the next portion of the personality test in order to be selected for a special task. All participants were then asked to come back for the second session the next week to discover whether they were selected. Participants who scored above the threshold on the validity tests were flagged as fakers and were randomly assigned to either a control or warning condition. The warning message held an identification and mild consequence component, while the control group was told that their data was lost due to a computer error and to complete the questionnaire again. All participants then completed a state affect measure for how they felt during the re-testing process. The results

showed greater accuracy of the retest scores than the control condition score comparing the motivated condition to the baseline personality scores.

This suggests that the warning message did deter individuals from continuing to fake. As for emotional reactions, state guilt was associated with greater accuracy of scores in the retest condition, whereas the control condition showed no relationship. State shame and state anger had no relationship with the accuracy of score, lending support to the hypothesis that guilt is the underlying emotion affecting score change.

In the second study business college students conducted the first personality assessment online as a baseline measure. The second portion of the study took part in a lab, where for the first half, participants were asked to analyze content commonly found on a resume and how it related to certain traits as a distractor task. After completing this first half, which was disguised as the main purpose of the study, participants were deceived into believing that the researchers were working with a company called Insat Corporation. They were asked to help with the development of a pre-screening measure for the company, and if they were interested in an internship position, their responses would be reviewed along with the program's needs. The participants took the first portion of the personality test and were then presented with either a control or retest condition. Those in the control condition were told that there was a computer error and to please complete the survey again, while the experimental group was given a warning message with an identification component and asked to answer the questions again in order to be considered. There was very little accuracy change from time 1 to time 2 for the control condition. For the experimental condition, the scores for those that were flagged as fakers improved in accuracy upon retesting. Those who were not flagged, however, decreased the accuracy of their scores upon retesting.

Most recently, the middle warning technique was examined for the potential negative effects, as well as the strength of the warning effect for varying levels of distortion (Burns, Fillipowski, Morris, & Shoda, 2015). In this study, the authors tested the behavioral effects of negatively, positively, and accusatorily worded warning messages. This study was the first to examine differentially worded messages with the middle warning design. The authors found that accusatory warning messages held the highest effect on score reduction out of the three warning styles. This supports the efficacy of detection and consequences components of warning messages put forth by Dwight and Donovan (2003). The authors additionally found that impression management moderated the effects of the warnings, in that the higher the level of faking in the initial test, the greater the reduction in their personality scores in the second post-warning test (Burns et al., 2015). In regards to the test-takers reactions to the various types of middle warning messages, the authors found that the positively worded warnings increased test-taker's motivation, but the negatively and accusatorily worded messages resulted in increased levels of anxiety (Burns et al., 2015). The measure of anxiety used in this study, however, was specifically test anxiety and not directed specifically towards the feelings during the warning message.

Current Study

Contributions

Unlike previous middle warning studies, the current study assessed multiple emotions in direct reaction to the middle warning message. The study design was based off of the work done by Ellingson et al. (2012) and Fan et al. (2012). This study combined the procedure put forth by Fan et al. with the examination of emotions evoked during the warning, as examined by Ellingson et al. (2012), in an attempt to further explore the internal mechanisms behind this score reduction. The current study aimed to more thoroughly investigate the emotional mechanisms of the middle warning through the measurement of physiological autonomic responses in conjunction with a self-report emotions survey. While Ellingson et al. (2012) identified guilt as a potential mechanism; I examined four emotions including guilt, fear, anxiety, and anger. Ellingson et al. used a self-report emotions survey at the end of the personality test. I am concerned that self-reported emotions can be influenced and biased by numerous factors such as impression management, fatigue, or simply not remembering the exact emotions felt during the target time. The addition of physiological measures will give us real-time unbiased data to decipher the elicited emotions, thus overcoming the limitations associated with self-reports. The current study addresses a recent call by several I/O scholars for moving away from relying on self-reports and exploring other alternative measures (e.g., Heaphy & Dutton, 2008).

Additionally, I was unable to locate any previous studies that examined this test re-test paradigm using the less traditional measure of the dark triad components of personality. The dark

triad was first introduced by Paulhus and Williams (2002) and describes aversive personalities that are still within normal functioning range. The three identified personality traits that compose the dark triad are Machiavellianism (cynical and manipulative), Psychopathy (impulsive and low empathy), and Narcissism (entitlement and grandiosity) (Furnham, Richards, & Paulhus, 2013). The Dark Triad has become a highly popular area of research in the I/O field recently as it relates highly to workplace behaviors such as toxic leadership and counterproductive work behaviors (Furnham, Richards, & Paulhus, 2013). Studies have also shown that the three dark measures of personality are highly related to self-monitoring behaviors (Rauthmann, 2011) and narcissists in particular engage in high levels of impression management (O'Boyle Jr, Forsyth, Banks, & McDaniel, 2012).

Therefore, this study is the first to examine if behavioral responding to warning messages on The Dark Triad measure behaves similarly to the more traditional Big Five measures of personality. If we are able to better understand what the affective response is to this middle warning, we can better inform practitioners if this is a viable way to manage applicant faking.

Study Design

In the current study, I randomly assigned a faking or honest condition and a warning or control message in the middle of the test, yielding four cells: (a) faking and warned, (b) honest and warned, (c) faking unwarned, and (d) honest unwarned. Every cell was presented with an initial test and a retest. This is contrary to the Fan et al. (2012) study design, which used an initial block of a few personality items and a main block of all personality items. The scripts for the warning and control message and the instructions for the honest and faking conditions can be found in the procedures section. The faking and honest instructions have been written in a manner suggesting the personality test is for job employment, only differing in the manner to

which they should respond. This is so that the warning message made sense to both honest and instructed faking participants.

An examination of the test-takers emotions took place in two parts. The first of these was an emotions survey at the end of the testing process asking them to recall their emotions at the time of the warning message. Due to the retrospective nature of this survey measure, these emotional reactions will be examined in terms of a between subjects design.

The second is a dynamic examination of the respondent's physiological reactions. As this is measured over time, these data were examined as a within subjects design. The physiological reactions from within subjects were then compared to the overall emotional reactions of the between subjects survey responses to examine for support and overlap. The following sections will elaborate on the survey emotions that were expected, and subsequently, the physiological reactions expected that would correspond with these emotions.

Emotions

Bradley and Lang (2000) define emotions as “action dispositions, mobilizing the body for behavior, but in which the overt action itself is often delayed or totally inhibited.” There are three dimensions to measuring emotional reactions. The first of these is arousal, which measures the amplitude of an emotion that ranges from calm to excited. The second is valence, which refers to the perception of the emotion being positive or negative. Finally emotions can be categorized into specific dimensions encompassing both their valence and arousal called discrete emotions (Bradley and Lang, 1994). Not all valenced reactions can be established as emotions, however. Moods are distinguished from emotions in that “moods do not have specific and stable motivational functions, but only informational function” (Kreibig, 2010). Moods tend to be brought on slowly and are more enduring. Emotions, however, are short-lived and brought on by

a sudden stimulus (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). Moods, in contrast to emotions are changes in feelings that occur without any influence from, or on, autonomic responses. It, therefore, becomes vital to distinguish which reaction will occur in the proposed study. Due to the nature of the warning component being sudden and unexpected, we can infer that the physiological response being measured at the moment of the warning message is, in fact, an emotion rather than a mood.

The response at the moment of the warning message is likely an emotion, but the next step is to examine how to classify that emotion. There have been countless theories regarding what emotions are primary and discrete, and what emotions are secondary or dimensional. Early researchers such as Watson (1924) argued that there could only be emotions based upon directly observable behavior prompting the discrete categories of fear, rage, and sexual performance. More modern taxonomies include a much broader range of classifications. Kemper (1987) argued that there were four primary emotions that paired with corresponding secondary emotions. The four primary emotions are anger, fear, depression and happiness, while the secondary corresponding emotions in order are shame, guilt, resignation, and pride as well as others. Paul Ekman (1992) argues for six distinct emotions such as anger, fear, sadness, enjoyment, disgust, and surprise with the possibility of contempt, shame, guilt, awe, and embarrassment as other distinct constructs.

To determine what the potential emotional reactions may be we must understand the context of how they are being elicited. Unlike a traditional warning message, which is given to everyone before testing and does not explicitly identify any individual as a faker, a middle warning message directly targets one individual of potentially faking after monitoring their behavior. Therefore, there are two conditions for which we must predict what emotional reaction

is likely to occur. The first is an individual is aware that they have been faking in the initial portion of the test when they receive the warning message, and therefore have been rightly targeted. The second is that the individual feels as though they have been giving honest responses throughout the initial testing, and subsequently feels as if they have been wrongly accused. These two scenarios should produce two patterns of emotional responses.

Cell 1 – warned fakers. The first scenario involves an individual who has been rightly accused of faking behavior. This test taker is aware that they were skewing their response patterns and believed that they could not, or hoped that they would not, get caught. Guilt is a discrete self-conscious emotion that arises from displaying behavior that is contrary to certain expectations (Tangney, 1990), and is social and interpersonal in nature, stemming from some type of differentiation from cultural ethics and norms (Barrett, 1995; Lewis, 2000; Tangney, 1999). Lying is a cultural faux pas, and therefore when an individual is responding dishonestly and this is confronted, such as with faking and the warning message, it becomes likely that the individual will express guilt. The guilt felt from having been dishonest will lead to those individuals subsequently reducing their scores (that is, responding more honestly) when given a second chance. Individuals in the warned non-faker condition have not been lying and should therefore not have the same guilt reaction to this confrontation. Therefore:

Hypothesis 1: Self-reported guilt will be significantly higher for individuals in the warned-faker condition than in any of the other three study cells.

While guilt is expected as the primary emotion for those who have been responding dishonestly and are shown the warning message, fear could potentially act as another mechanism. Kemper (1987) proposed that emotions such as fear stem “from interaction outcomes where actors are subject to the power of others because that power is greater than their

own,” due to the detection and consequences portion of the warning message, and an individual may feel fear because they are subject to the decisions of the test administrators in terms of personality score acceptance. An applicant could potentially become fearful of those consequences that come with continuing to fake, and subsequently reduce their personality scores because they feel threatened. This is in contrast to the idea that individuals will lower their scores due to moral reactivity from the warning itself. Those who are warned non-fakers may feel fear, but this emotion will be overwhelmed by another emotion to be explained below in cell 2. Kemper (1987) also argued that there is a very strong association between guilt and fear. Guilt is the learning outcome from the threat of punishment from a parental figure and therefore, guilt is fear of punishment from engaging in prohibited actions (Kemper, 1987). Similarly, Handler and Honts (2007) stated that, “the fear of punishment (or fear of consequences) theory postulates a guilty examinee will experience autonomic arousal as a result of fear of consequences of discovery or false accusation” supporting the notion that fear may be a dominant emotion in this cell. Therefore:

Hypothesis 2: Self-reported fear will be significantly higher for individuals in the warned-faker condition than in any of the other three study cells.

Cell 2 – warned non-fakers. While guilt and fear are the proposed mechanisms behind those who fake and are given the warning component, the question arises as to what will be the mechanisms behind those who do not fake yet still receive the warning message? Kemper (1987) states that anger stems from “interaction outcomes in which expected, customary, or deserved status has been denied or withdrawn by another actor who is seen to be responsible for the reduced status.” Anger is a result of someone else’s actions that threaten an individual’s desires unjustly. When an individual is told that their expected outcome of doing well on the test may be

taken away due to an unfounded claim of dishonesty, that individual may express a negative reaction of anger. Those individuals who were performing honestly on the test, with the desire to gain employment from the potential internship opportunity, will feel anger when they feel they are wronged and perceive that they must reduce their scores in order to avoid negative consequences. This score reduction denies them of what they feel is their deserved status as the result of another actor.

Hypothesis 3: Self-reported anger will be significantly higher for individuals in the warned-honest condition than in any of the other three study cells.

While the initial emotions felt will be anger, I posit that the actual emotion that acts as the mechanism for score reduction will be a fear of the repercussions for not doing so. The test-taker who has been responding honestly in the initial testing block will initially feel angry when accused of providing false information, but will then feel that they must change their answers for fear of being selected out of the testing process. This fear reaction, however, may be dulled by the initial anger reaction. Therefore:

Hypothesis 4: Self-reported fear will be significantly higher for individuals in the warned-honest condition than in either control condition, but not the warned-faker condition.

Cells 3 & 4: Unwarned fakers and unwarned non-fakers. Cells 3 and 4 consist of test takers who have faked, yet have received a control message, and test-takers who have not faked, and have received a control message, respectively. The control message is designed to be a neutral stimulus and should therefore, in theory, not elicit any type of emotional response. The control message states that this is a random system check and should therefore have no effect on the test-takers guilt, fear, or anger. Therefore:

Hypothesis 5: Self-reported guilt, fear, and anger will be significantly lower in the control condition, than in the warning condition.

However, this is a somewhat stressful test-taking environment and any type of pop-up message on the computer screen could be met with a certain level of anxiety or frustration. There is also the possibility that this reaction of anxiety could be true of any of the four study cells. Rather than any of the previously proposed emotional reactions (fear, guilt, and anger), test-takers could simply be experiencing general state anxiety. A measure of anxiety, therefore, will be included in the emotions survey and examined along with the physiological responses.

Physiology

The association between emotions and autonomic responses was first suggested by William James (1884). Since that time, there has been much debate regarding the validity of classifying emotions through the measurement of physiological responses. James (1884) argued that various autonomic responses produce varying emotions, while Cannon (1927) argued for the opposite trajectory, where varying emotions produced the varying autonomic responses. Others have argued that it is neither, but rather a combination of the two with both theories being relatively correct (Schachter & Singer, 1962). Regardless of the directionality, there is relative consensus that autonomic responses and emotions are associated with one another. Cacioppo, Berntson, Larsen, Poehlmann, and Ito (2000) argue that there is no question of “whether emotion-specific autonomic patterns occur, but under what conditions such patterns occur.” Barrett (2012) argues that there is a multitude of evidence indicating that each emotion category has a unified biological basis. She defines emotions as ontologically subjective categories created by humans to make meaning of physical events and to prescribe actions (Barrett, 2012). Emotions allow us a quick cognitive processing of outside stimuli in order to prompt biological

reactions for reactive behavior. While the connection between emotions and physiological reactions has been widely accepted, the exact relationship between discrete emotions and specific responses is less clear. Despite the broad wealth of information related to the psychophysiology of emotion, there has yet to be consensus regarding the specific physiological pattern for each specific emotion. For example fear can induce either a fight, flight, or freeze response; therefore context becomes imperative to understanding and interpreting the psychophysiology of emotions (Bradley & Lang, 2000).

Borrowing from the Funkenstein hypothesis (Funkenstein, 1955), Kemper (1987) argues that the two emotions fear and anger can be differentiated physiologically due to the various neurotransmitters involved in eliciting these emotions. The release of the neurotransmitter epinephrine is associated with fear, while the release of norepinephrine is associated with anger. Both neurotransmitters activate the sympathetic nervous system but in varying ways. A study by Ax (1953) sought to differentiate the specific physiological reactions between the two constructs. The results showed that for anger, there was a greater average reaction for heart rate falls, the number of galvanic skin responses, and muscle tension increases. For fear there was a greater average for skin conductance increases, number of muscle tension peaks, and increase in respiration rate. However, other scholars argue that there is no way to confidently claim differential autonomic responding among various emotions, as there have been inconsistent patterns in the literature. Some claim this is due to the presence of moderator variables (Cacioppo et al., 2000), while others state that it is because of the impossibility of distinguishing emotions through autonomic responding (Barret, 2006).

In an attempt to address this issue, a recent review of autonomic responding was conducted by Kreibig (2010); aiming to consolidate the vast literature and define a clear

differentiation among the discrete emotions. She compared 134 publications examining autonomic responses for various discrete emotions and presented the results of the most consistent patterns found for 22 various emotional reactions. Kreibig (2010) compared a number of physiological measures in regards to emotions, but for brevity, only those most relevant to the postulated emotional mechanisms will be discussed.

One of the first physiological responses that can aid in identifying emotional response patterns are beats per minute (BPM) or heart rate. Heart rate is dependent on autonomic neural regulation, and is controlled by the balancing act between both the parasympathetic nervous system (PNS) and sympathetic nervous system (SNS; Acharya, Joseph, Kannathal, Lim, & Suri, 2006). The PNS and SNS are the two parts of the autonomic nervous (ANS) system, which serves as the, “regulator, activator, coordinator, and communicator,” of the body’s complex, “network of nerves, organs, and biological sensors” (Levenson, 2014, pp. 101). In terms of cardiology, the SNS prepares the body for physical action, while the PNS is responsible for slowing and resting the heart rate (Levenson, 2014, pp. 102). Increases in heart rate can represent an increase in SNS and subsequent decrease in PNS, while a decrease in heart rate may represent the reverse (Acharya et al., 2006). In terms of the emotional mechanisms underlying the middle warning message, according to Kreibig (2010) we should see an increase in BPM for those experiencing guilt, but a decrease for those experiencing either fear or anger. However, as BPM is controlled by the cardiac sinoatrial node, which is subject to both subsystems of the ANS, it is more appropriate to examine heart rate variability to assist in determining whether the SNS or PNS is being activated (Berntson, Bigger, Eckberg, Grossman, Kaufmann, Malik, Nagaraja, Porges, Saul, Stone, & Van der Molen, 1997).

Within heart rate variability (HRV) there are multiple categories regarding level and frequency. In regards to frequency, it can be categorized as high or low and within those either very high or very low. This results in four classifications of heart rate variability, which are: low frequency (LF), very low frequency (VLF), high frequency (HF), and very high frequency (VHF). LF and VLF have traditionally been interpreted as representing sympathetic cardiac control. The sympathetic nervous system is associated with LF is said to represent “oscillations related to regulation of blood pressure and vasomotor tone” while VLF is understood to relate to “thermoregulation and kidney functioning” (Reyes del Paso, Langewitz, Mulder, Roon, & Duschek, 2013). There have been recent findings, however, to suggest that these measures of heart rate variability are not quite as straightforward to interpret as once thought. Reyes del Paso et al. (2013) argue that LF is actually influenced by the parasympathetic nervous system, while Billman (2013) argues that it is a convoluted mix of both sympathetic and parasympathetic influences. Similarly, the review article by Kreibig (2010) only displays an expected pattern of results for the emotion fear, in which both measures increased from baseline. Therefore, LF and VLF should be interpreted in terms of its additional support to patterns formed by other, more straightforward, measures.

The literature regarding high frequency (HF) and very high frequency (VHF) is much more consistent, and it is widely held that these components of HRV stem from a vagal origin, and are therefore representative of cardiac parasympathetic tone and the effects of respiration on heart rate (Reyes del Paso et al., 2013). HF and VHF will be analyzed in terms of an increase or decrease from baseline, which, when analyzed as a pattern with the other measures, will lend support to either an increase or decrease in parasympathetic activity similarly to LF and VLF.

Once again, an expected pattern of results was only found for the emotion fear, in which HF and VHF increased from baseline.

Continuing with measures regarding the heart, respiratory sinus arrhythmia (RSA) is reflective of “tonic and phasic vagal influences on the heart” (Overbeek, van Boxtel, & Westerink, 2012). There has been substantial variability in the emotions literature regarding RSA and it has been suggested that this could be based on the induction method, situation, as well as various other inconsistencies across studies (Overbeek, van Boxtel, & Westerink, 2012). Therefore, RSA can similarly be analyzed in terms of its contribution to other measures of physiological responding and interpreted according to patterns that arise which are consistent with those found in Kreibig’s (2010) review article. RSA’s main contribution to distinguishing emotions in regards to the middle warning message is an increase from baseline when experiencing anger, as postulated with warned non-fakers.

Pre-ejection period (PEP) is another measure regarding the heart and is defined as the “interval from the onset of the ECG Q-wave to the onset of left-ventricular ejection” (Allen, Fahrenberg, Kelsey, Lovallo, & Doornen, 1990). It is inversely related to myocardial contractility and therefore relates to sympathetic influences on the heart (Newlin & Levenson, 1979). This measure, similar to the others, can be interpreted in accordance to its contribution to the overall pattern of autonomic responding. In all three relevant emotional responses, PEP decreased from its baseline measure, so while it cannot be interpreted in terms of differentiation, it can be useful in lending overall support.

A similar physiological measure, aiding in the overall identification of emotional responses, is stroke volume. Stroke volume (SV) is the amount of blood pumped with each

heartbeat from the left ventricle (Berntson, Quigley, & Lozano, 2007). A decrease in SV has been found in a number of studies that utilize pictures to elicit fear of threat responses (Dimberg, 1986; Codispoti & De Cesarei, 2007; Bernat, Patrick, Benning, & Tellegen, 2006), which provides support for the proposed emotional mechanism regarding both warned fakers and non-fakers. Therefore a decrease in SV could indicate a fear response when receiving the warning message.

Cardiac output (CO) is directly related to stroke volume and is “the amount of blood pumped by the left ventricle into the aorta per unit of time” (Berntson, Quigley, & Lozano, 2007, pp.186). This measure of the cardiovascular system involves oxygen transport from the heart throughout the body (Pinsky, 2003). An increase in cardiac output, and therefore an increase in oxygen flow to the extremities, indicates readiness for survival and an increase in the SNS. However, according to the review article by Kreibig (2010), in order to differentiate between the three proposed emotional mechanisms, there should be no change for CO if the reaction is guilt, an increase in CO for fear, and a decrease in CO if the emotion is anger directed away from the self. Presumably the anger emotion would be directed toward the test administrator or organization as a whole, as this mechanism is associated with non-fakers who feel they have been wrongly identified; supporting this pattern. However, if the individual feels anger directed inward, perhaps due to a small amount of faking (as the cutoff scores are arbitrary) the pattern of CO would increase from baseline, making it indistinguishable from fear.

The last measure that is potentially useful to lend support in identifying the emotional mechanisms are skin conductance responses (SCR). SCRs represent changes in eccrine sweat gland activity stemming from the sympathetic nervous system (Khalifa, Isabelle, Jean-Pierre, & Manon, 2002). This measure will provide useful data in regards to interpreting results, as it is

solely influenced by the SNS. The results from this measure will provide direct insight into whether the SNS or PNS is being activated. In Kreibig's (2010) article SCRs increased from the baseline for fear and anger. Clearly, there is a large amount of overlap in physiological responses for each emotion, necessitating identification of a large pattern to discern one emotion from another.

Overall the pattern of physiological responses are complex regarding identifying individual emotions, but when paired with self-report data and examined at various points in time, they can lend invaluable insight into the warning message process. For clarity and ease, Table 1 summarizes the expected pattern of physiological response changes from baseline to the warning message for each of the proposed emotional mechanisms. An upwards-facing arrow represents an increase from the baseline, a downward facing arrow a decrease from the baseline, and a dash is no change from the baseline. N/A indicates there was no available information regarding that physiological measure in regards to that emotion.

Cell 1 – warned fakers. As a reminder, I argue that in Cell 1 two emotions might be evoked: guilt and fear. In our specific paradigm, the threatening fear of punishment may be particularly salient. Greenwald, Cook, and Lang (1989) measured physiological reactions to multiple emotion-eliciting pictures, including fear, and found that increased heart rate was significantly related to more pleasurable judgments, while decreased heart rate was related to unpleasant judgments. This direction can be reversed when the subject prepares for a threat, and engages the fight or flight response, but as the fear emotion being elicited in our study is not a physical threat, we expect a decrease in heart rate associated with the unpleasant judgment (Greenwald, Cook, & Lang, 1989). Similarly, a study conducted by Bernat, Patrick, Benning,

and Tellegen (2006) found that there was an increase in skin conductance responses and a decrease in heart rate in reaction to imagery of threatening figures. In a comparison of real-life versus imagined elicitations of fear and anger, Stemmler, Heldmann, Pauls, and Scherer (2001) found that for real-life induction of fear there was a decrease in PEP, a decrease in stroke volume, an increase in CO, and an increase in SCRs. In addition to these patterns a meta-analysis by Kreibig (2010) similarly found a decrease in stroke volume as well as an increase in heart rate variability.

Therefore, for fear we expect a physiological pattern of a decrease in heart rate and stroke volume, and an increase in skin conductance, low frequency, and high frequency heart rate variability responses between the physiological state during the baseline, while the body is at rest, and the physiological state during the presentation of the warning or control message.

The other possible emotion that could arise within Cell 1 is guilt. Elaad (2009) performed a study in which the level of state guilt of the participants was manipulated through providing varying levels of information regarding a crime. The author found that skin conductance responses were highest for those who had been in the guilty condition. Another study conducted by Fourie, Rauch, Morgan, Ellis, Jordaan, and Thomas (2011) manipulated guilt by implementing a deception in which that participant was given more money than they were supposed to from the research assistant, asked to lie about it, and then confronted by the research supervisor. This study found a similar increase in skin conductance, as well as an increase in heart rate, and a decrease in pre-ejection period. If the emotion felt is guilt, the increase in BPM and SCR are straightforward indicators of an increase in SNS activity, and preparing the individual to act. However, according to the review article by Kreibig (2010), in order to differentiate between guilt and other emotions, there should be no change for cardiac output

Therefore, we should see the following pattern of physiological responses between the physiological state during the baseline and the physiological state during the presentation of the warning or control message: Increased heart rate, increased skin conductance, decrease in PEP, and no change in cardiac output.

Cell 2 – warned non-fakers. I suggested that in Cell 2 two emotions might be evoked: Anger and fear. As previously mentioned, Stemmler et al. (2001) elicited real-life anger and fear responses and found that for real life anger, there was a decrease in PEP, a decrease in heart rate, a decrease in stroke volume, and an increase in cardiac output and SCRs. Fear and anger have been shown empirically to be similar both theoretically and physiologically (Carver, & Harmon-Jones, 2009). However to differentiate these two emotions from a physiological standpoint, we will rely on the measure of respiratory sinus arrhythmia, as well as self-report. Kreibig (2010) found that there is an increase in RSA in conjunction with a decrease in heart rate when anger is elicited.

Therefore, anger should manifest itself in a decrease in heart rate, PEP, and SV, and an increase in skin conductance, CO, and RSA between the physiological state during the baseline and the physiological state during the presentation of the warning or control message.

Although the emotion felt at this moment should primarily be anger, there is still the underlying emotion of fear and threat at potentially being taken out of the selection process. Therefore we may see a physiological response pattern similar to the one hypothesized in cell 1. Self-reported emotions will assist in clarifying the dominant emotion within this cell.

Cells 3 & 4: Unwarned fakers and unwarned non-fakers. The control message is designed to be a neutral stimulus and should therefore, in theory, not elicit any type of physiological response. However, as suggested there could be a certain level of anxiety or

frustration. *Therefore the physiological response change should either be none, or similar to a pattern of anxiety which includes an increase in heart rate, skin conductance, and cardiac output, but no change in SV between the baseline and the presentation of the control message.*

All physiological responses will be initially examined in terms of their change from the traditional baseline, or resting period, to the warning message. However, there is a question that perhaps the instructions informing individual's to respond a certain way on a measure often used for job selection, may have a significant effect on their subsequent physiology. Therefore, physiological responses for both the baseline and the pre-test will be compared. If there is a significant difference, there will be further exploration into what response patterns will be found if using the pre-test as a baseline measure of physiological responding.

Table 1: Proposed pattern of physiological responses.

		HR	SCR	CO	RSA	LF/VLF	HF/VHF	PEP	SV
Cell 1 Warned Faker	Guilt	↑	↑	=	N/A	N/A	N/A	↓	N/A
	Fear	↓	↑	↑	N/A	↑	↑	↓	↓
Cell 2 Warned Honest	Anger	↓	↑	↑	↑	N/A	N/A	↓	↓
	Fear	↓	↑	↑	N/A	↑	↑	↓	↓
Cell 3 & 4 Unwarned faker/ honest	No emotion or anxiety	= / ↑	= / ↑	= / ↑	N/A	= / ↑	= / ↑		=

Note. HR= heart rate, SCR= skin conductance response, CO= cardiac output, RSA= respiratory sinus arrhythmia, LF/VLF= low/very low frequency, HF/VHF= high/very high frequency, SV= stroke volume.

Methods

Participants

The sample consisted of 244 undergraduates at Auburn University. The 2x2 design yielded 4 separate cells: (a) faking and warned (n=59), (b) honest and warned (n=63), (c) faking unwarned (n=63), and (d) honest unwarned (n=59). For testing the over compensatory response, mass screening data was only found for 215 students: (a) faking and warned (n=53), (b) honest and warned (n=55), (c) faking unwarned (n=55), and (d) honest unwarned (n=52). Participants were given extra credit and entered into a random cash drawing for their participation.

Procedure

Participants were recruited through the SONA web system. Participants were required to complete a mass screening questionnaire before they could participant in this lab study, which included the Dark Triad, School-Specific Conscientiousness, and Impression Management. The first step once they entered the lab was to introduce the participants to the physiological portion of the study:

Physiological Introduction Script

“Hello, I am _____. Thanks for coming to the lab and completing our study. We are interested in examining the personality and physiological responses of individuals as they apply for a job position. Therefore, with your consent, you will be attached to physiological recording equipment. There may be minor discomfort when removing the

adhesives from the electrodes after the test is finished, similar to lifting tape off the skin. Other than this, there should be no other discomfort, but please inform us if at any point you feel discomfort. Please read over the consent form carefully and decide if you would like to proceed with the study.”

After the introduction, consent forms were distributed to the participants. Once the participants gave their consent, the test-takers were attached to the physiological recording equipment. The next step was to ask participants to sit quietly resting for two minutes to obtain the individuals resting physiological measurements.

Resting Period Script

“In order to collect a resting baseline of your physiological measurements, we need to run the data collection while you are relaxed. This will only take two minutes, so please sit quietly in a resting state.”

After the two-minute resting period, participants were introduced to the study.

Survey Introduction Script

“As previously stated, our lab is interested in examining the personality and physiological responses of individuals as they apply for a job position. Please thoroughly read your instructions as to how to respond to the personality questionnaire, as they may be different than instructions you have received on previous personality tests. The personality test should take no more than 30 minutes so please take your time and respond carefully.”

The participants then began the testing process and were randomly assigned to either a fake-good or honest condition by the computer. Therefore, in this study the faking versus honest condition

was directly manipulated rather than using a measure of faking to determine fakers and non-fakers. There are multiple benefits to this approach as it allows a more even distribution of participants across the study cells, and it has the potential to intensify the emotional reactions of the participants to the randomly assigned warning. The introductions for these conditions were adapted from van Hooft and Born (2012).

Honest introduction

“Thank you for participating in this study, which includes a random drawing at the end of the semester for one of ten \$25 cash prizes. In the next screens you will be presented with a personality questionnaire with five response options. This personality questionnaire has been routinely used to select college students into entry-level positions by many organizations. **We ask that you answer the questions as honestly as possible.** Your answers will remain completely confidential and anonymous, and will be used for research purposes only. For this study we are interested in who you really are. Therefore it is very important that you answer the following questions as accurately and honestly as you can.”

Fake-good Introduction

“Please imagine that you are graduated and are applying for a job. As part of the selection procedure you are presented with the following personality questions with five response options. **Please answer the questions such that you will come across as the ideal employee.** For this study we are not interested in what your real answers for each question would be. Instead, for each question please select the answer that you feel will

give you the best rank and make you look like the most suitable job applicant. The top ten participants with the most ideal scores will receive \$25.”

The \$25 bonus was to add an element of motivation to those in the faking condition. This, however, is a slight deception, as the 10 \$25 bonuses will be given out through random drawing in order to ensure fairness for those randomly sorted into the honest condition. (Participants were debriefed on this deception at the end.)

Participants then began the testing process, which included an initial test of school-specific conscientiousness personality items, the “Dirty Dozen” dark triad personality measure, and an impression management measure. The impression management measure was included to provide a validity check on the fake-good and honest instructions. After the initial test, participants were randomly assigned to either the warning or control condition. Participants received the detection and consequences warning or a control, which stated that the interruption was simply a random system check. Both conditions were setup to display for 45 seconds each, in order to control for a potential length of interruption confound. The warning component included the traditional identification and consequences component. The first reason being that this method has shown validity above other styles of warning messages (Dullaghan, 2010) and the second is that due to its more harsh tone, it is more likely to evoke the emotional reaction attempting to be discerned from the warning process.

The Warning Message

“Warning Warning Warning”

Thank you for participating in this survey. However, we have noticed some *unusual response patterns* in your answers and wish to clarify the issue. The personality inventory, which you are completing, has embedded social desirability scales. These scales identify people who might have tailored their responses to what they believe others would like to hear.

Your response profile up to this point is similar to that of someone who is known to be answering in a socially desirable way. We do not intend to insult your integrity; we only want to get a clear understanding of who you are.

Thus, we would like to give you the opportunity to re-take the survey. Remember, be yourself and answer each question as it best describes you. Finally, rest assured that your previous responses on these inventories will be discarded. However, we have found in the past that some participants had repeatedly distorted their response. These individuals were quickly discovered and were immediately disqualified from receiving the cash bonus.”

Control Message

“Thank you for participating in this portion of the selection process. A random system check indicates the testing system is working well. Please continue the test. Be reminded that as part of the testing procedure, some of the items will be presented twice. So don’t be surprised if you see some of the items showing up again on the screen.”

After the warning/control message, all participants were then guided through the retest, which contained all of the same measures as the initial test. After the participants completed the

second test, the participants were asked to complete a brief emotions survey reflecting their state emotion at the moment of the warning/control.

Pre-emotion Questionnaire Script

“You will have noticed that you received an interruption in your testing process in the form of a system message through a pop-up as seen above. You will next be asked to complete an emotions survey regarding how you felt in that moment that you received that interruption. Try to recall your emotions as accurately as possible and please respond honestly.”

The emotions survey included measures for all of the hypothesized emotions as well as irrelevant emotions so as not to skew responses (See Appendix A). They were then debriefed and thanked for their participation (See Appendix B for script). At the end of the semester, ten participants were randomly selected to receive \$25.

Measures

School-specific Conscientiousness (SSC; Schmit, Ryan, Stierwalt, & Powell, 1995):

The SSC is a 24-item measure used to determine an individual’s score on three dimensions of conscientiousness: achievement striving, competitiveness, and self-discipline.

Dark Triad (The Dirty Dozen; Jonason, & Webster, 2010): The “Dirty Dozen” dark triad is a 12-item measure of three traits: Machiavellianism, psychopathy, and narcissism.

Impression Management (Bing et al. (2011); Paulhus (1988)): 10 impression management items will be included in both the initial and main blocks and are interspersed

within the personality questions. 7 of the impression management items are from Bing et al. (2011) and the other 3 were chosen from Paulhus's (1988) BIDR-6.

The Positive and Negative Affect Scale-X (Watson & Clark, 1999): A self-reported emotions survey will be included at the end of the study to provide supporting evidence for the results of the physiological analysis. The scale consists of various words that describe different feelings, which correspond to the basic emotions of: fear, hostility, guilt, sadness, joviality, self-assurance, and attentiveness. Example words for the fear scale would be “frightened” or “nervous,” while an example of joviality would be “happy” or “joyful”. Participants rate their level of arousal when seeing either the control or warning message on a 5-point Likert scale from “very slight or not at all” to “extremely” for each word.

The Personal Feelings Questionnaire (Harder & Zalma, 1990): A self-reported state guilt and state shame scale. This scale will be presented with the PANAS-X on the same 5-point Likert scale and included items such as “embarrassed” or “intense guilt” for the guilt sub-scale and “feeling stupid” or “self-conscious” for the shame sub-scale.

Fear of Punishment (Fan et al., 2015): A self-reported survey regarding the extent to which test-takers feel they have to change their answers. Participants will be asked to rate on a 5-point Likert scale the extent to which they agree with statements such as, “The system message I received in the middle of the test made me very concerned about possibly failing the psychological test.”

State-Trait Anxiety Inventory (Marteau & Bekker, 1992): A self-reported state anxiety scale. Respondents will be asked to rate on a 4-point likert scale the extent to which they agree with items such as, “ I am worried” and “ I feel content”. The tenses will be changed to

past tense, since respondents will be recalling their state emotion at the moment of the manipulation.

Skin Conductance (Biopac product #EDA100C-MRI): Two electrodes will be placed on the third and fourth finger and measure electro-dermal response. This measure will digitally record sympathetic arousal onto an hp computer.

Heart Rate Variability (Biopac product #EMG100C-MRI): Two electrodes will be used to measure heart rate. One electrode will be placed on the right collarbone, and the other on the participants left rib. The measure will digitally record parasympathetic activity.

Impedance Cardiography (Biopac product #NICO100C-MRI): Two electrodes will be placed on the back of the neck, and two on the lower back of the participants as a measure of sympathetic arousal.

Respiration (Biopac product #RSP100C): Respiration will be measured with a respiration belt wrapped around the participants' upper waist. The measure will offer greater insight into heart rate variability.

Results

Analytic Strategies

All hypotheses were analyzed using independent t-tests in which the file was split between warning vs. control, and honest vs. instructed faker. The four resulting means were compared for each emotional reaction. The two manipulations were dichotomously scored in that honest=0 and faker=1 for the variable instructions, and control=0 and warning=1 for the middle message variable. All personality and the impression management measure were calculated using a composite score for time 1 and time 2. The Dark Triad scale was reverse scored for ease of interpreting the results such that, an increase in score indicated an increase in socially desirable responses. The reduction of scores from time 1 to time 2 was accounted for by using time 2 scores as a dependent variable with time 1 scores entered in as a covariate. All mediations and moderated mediations were tested using the PROCESS macro for SPSS (Hayes, 2013).

Physiological analyses were performed using Acqknowledge Software 4.4. Focus areas were created for each portion of the study resulting in the following sections: instructions, baseline, initial block, manipulation, main block, emotions, and debriefing. One person created the focus areas for all participants in the study to ensure consistency.

LF, VLF, HF, VHF, and RSA were analyzed using the multi-epoch HRV and RSA-Spectral analysis. All of the standard pre-programmed settings were used with this analysis and

all measurement graphs were checked to ensure there was a properly recorded, visible heartbeat. Participant data, in which there was no decipherable heartbeat, were excluded from analysis.

HR, CO, SV, and PEP were calculated using the impedance cardiography ICG analysis. All of the standard settings for this analysis were kept except for the stroke volume method, in which the Kubicek/Rho method was used. This is because the height and weight of the individuals was not gathered, and results using this analysis do not rely on those measures (Height/Weight entries for CO using Kubicek/Rho method | BIOPAC.). The program will not run without some entry in the height and weight columns, however, so a standard height of 5 feet 7 inches, and a weight of 155 were used. Similar to the previous analyses, measurement graphs in which there was no discernable waveform were excluded from analysis. Participant's data were determined to be included based upon the clear distinction of having a properly recorded waveform versus having no recording at all, or a waveform that wasn't able to be interpreted resulting from improperly placed electrodes.

Lastly, to analyze skin conductance, the waveform was resampled from 1K Hz to 15.67 Hz on the EDA channel, as a frequency above 10Hz is recommended (Fowles, Christie, Edelberg, Grings, Lykken, & Venables, 1981). The number of skin conductance responses was then counted based upon a 45 second interval. The warning and control messages were standardized at 45 seconds across every participant so all skin conductance responses were used for that focus area. For the baseline the last 45 seconds of the baseline were used for comparison. However, as the results will show, there was no discernable data for skin conductance.

The output for all of the physiological analyses gave averages for each measure in each of the pre-selected focus areas. These averages were used for further analyses in SPSS to

compare the averages of the different focus areas. Each physiological measure was checked for outliers using the explore feature to produce boxplots. Participant's individual averages that were marked as outliers during both the baseline and the middle message manipulation were filtered out and excluded from further analysis. For all physiological analyses, a paired samples t-test was used to evaluate any group differences from the baseline to the manipulation.

Manipulation Check

Participants were randomly assigned one of two sets of instructions. One set of instructions asked them to respond to the survey with complete honesty, and the other set asked them to respond in a way as to make them appear like the ideal applicant. To determine if participants responded in the appropriate way, a measure of faking (IM) was included in the personality survey. At time 1, before the warning message was administered, participants in the honest condition had a mean impression management score of 44.67, while those with the faking instructions had a mean of 54.20. The instructed fakers had a significantly higher mean ($F(1, 242) = 77.1, p = .000$) than those who were instructed to be honest, suggesting that the instructions manipulation was successful.

The next manipulation check examines if the warning component worked successfully in reducing scores for those who were instructed to fake. Table 2 shows the mean score differences from time 1 to time 2 for all personality scales. For ease of interpretation, all Dark triad personality scales were recoded so that a higher score on any scale indicates a more socially desirable response. The results are fairly consistent with what was expected.

Table 2: Mean differences between time 1 and time 2 personality scales.

	Honest Control ^a		Honest Warned ^b		Faking Control ^c		Faking Warned ^d	
	M	SD	M	SD	M	SD	M	SD
Conscientiousness	.95	4.70	** -6.67	10.82	* -2.56	7.82	** -13.03	18.30
IM	.25	2.11	** -3.02	4.95	** -2.40	6.65	** -7.00	9.76
Machiavellianism	.25	2.20	** -1.62	2.84	.06	3.24	** -1.73	4.11
Psychopathy	-.12	1.35	** -1.38	2.87	-.35	2.14	** -1.75	3.31
Narcissism	.08	1.39	-.32	2.66	.03	2.74	* -1.29	3.87

Note. ^an=59. ^bn=63. ^cn=62. ^dn=59. A positive number indicates more desirable responses in time 2. A negative value indicates a reduction in desirable responses from time 1 to time 2 †p<.10, *p<.05, **p<.01

Individuals who were given the honest construction set and received the control message did not change their scores from time 1 to time 2 for any of the personality scales. Those who were asked to be honest and given the warning message significantly reduced their scores to be less desirable on all scales except narcissism. For those who were instructed to fake and given the control message, participants did not significantly change their responses on the dark triad measure, but did significantly reduce their scores on conscientiousness and impression management. Lastly, those who were instructed to fake and were given the warning message significantly reduced their scores on all five measures. The responses in the honest control, and faking warned behaved exactly as expected, in that the control message should not affect scores, and the warning message should significantly affect scores. There were a few results that did not fit this pattern and reasons for this will be reviewed in the discussion session.

Over-compensatory Response Check

When administering a warning message and giving participants a chance to re-take the test, there is a risk that instead of fakers correcting their scores to be closer to their true personality score, they will overcompensate. To test for this, all participants were required to have completed an at-home mass screening personality questionnaire, which includes all of the scales utilized in this study, in order to be eligible to sign up for this lab study. The participants

take the mass screening at home at their convenience and in theory has no motivational context to fake. This allows for capture of their true personality scores. In order to test if there is a compensatory response, the personality scores from that mass screening can be compared directly to the participants' personality scores at time 1, before the warning, and time 2, after the warning. The means of the personality tests at these three separate time-points can be found in table 3. As can be seen from a brief glance at the means of each scale, the scores for time 2 are consistently lower and closer to the mass screening scores for the positive personality scales, and higher and closer to the mass screening scores for the negative dark triad items. This suggests that there was no over-compensatory response, as none of the means in time 2 are lower than those in the mass screening.

Table 3: Means, standard deviations, and reliabilities for personality scales.

	Mass Screening ^a		Time 1 ^b			Time 2 ^c		
	(True Score)		(Initial Test)		α	(Retest)		
	M	SD	M	SD	α	M	SD	α
Conscientiousness	101.16	8.66	140.28	19.39	.94	134.58	21.35	.94
IM	44.48	4.85	49.87	9.84	.80	46.68	9.92	.79
Machiavellianism	15.40	5.18	16.80	5.09	.81	16.04	5.28	.83
Psychopathy	19.14	4.80	20.22	3.94	.80	19.32	4.56	.84
Narcissism	10.87	5.00	10.81	5.07	.79	10.44	5.19	.80

Note. ^an=215. ^bn=243. ^cn=243. IM= Impression Management. A larger score indicates more desirable responses.

The differences in scores between these three time-points were examined based upon both instruction sets and warning vs. control. First the file was split based on faker vs. honest instruction set and a paired samples t-test was conducted to examine any differences in the mass screening and time 1. The results can be found in table 4.

Table 4: Mean differences between mass screening and time 1 personality scores.

	Honest Instructions ^a		Faking Instructions ^b	
	M	SD	M	SD
Conscientiousness	**29.55	15.01	**48.60	18.40
Impression Management	.748	8.63	**9.97	10.56
Machiavellianism	** -1.22	3.43	**4.39	6.18
Psychopathy	** -.81	2.55	**3.19	5.30
Narcissism	** -1.45	3.11	**1.24	4.56

Note. ^an=107. ^bn=108. A positive number indicates an increase in score from the mass screening to time 1, or in other words, responding in a more desirable manner. †p<.10, *p<.05, **p<.01

These results show that with the exception of impression management in the honest condition, participants significantly increased their scores on the positive scales (conscientiousness and impression management) and significantly decreased their scores on the scales with the negative dark triad personality traits. Therefore individuals given the honest instructions claimed more socially desirable traits on conscientiousness, but responded less desirably than they reported in the mass screening for the dark triad traits. This suggests that individuals have different response behaviors for differently framed personality traits.

Participants who received the faking instructions claimed more socially desirable traits than their true personality scores for all of the personality measures, as expected. The next step is to examine if the participants personality scores more closely resemble their true personality scores after the warning message is implemented.

Table 5: Mean differences for mass screening and time 2 personality scores.

	Control Message ^a		Warning Message ^b	
	M	SD	M	SD
Conscientiousness	**37.27	19.59	**29.61	21.45
Impression Management	**3.86	10.68	.55	10.71
Machiavellianism	**1.73	5.83	-.18	5.45
Psychopathy	*1.07	4.72	-.56	4.67
Narcissism	.08	4.26	*-.98	4.33

Note. ^an=107. ^bn=108. A positive number indicates a higher score in time 2. †p<.10, *p<.05, **p<.01

As can be seen in table 5, with the exception of narcissism, those who received the control message still had significantly more socially desirable responses to the time 2 retest, than their true scores. This is to be expected, as they were not warned against socially desirable responding. For those who were warned, time 2 impression management, Machiavellianism, and psychopathy, did not show significant differences in scores from the participants' true scores. This suggests that after the warning message, participants report more honest personality scores for those three measures. The exceptions were conscientiousness and narcissism. Even after the warning message, participants reported significantly higher scores on conscientiousness than their true score. This may be due to the context of taking the test at home versus, in a school setting where school specific conscientiousness items would be more salient. The participants, however, did reduce their conscientiousness from time 1 to time 2 overall as can be seen in table 3.

Narcissism had a significantly lower score in the retest, than in the mass screening, and the conclusions drawn are contrary to conscientiousness and IM. Paired with the fact that all of the dark triad retest scores were higher than those in the mass screening (only narcissism was significant), it could suggest that perhaps individuals are either over compensating on these scales or underreporting their negative traits during the mass screening. This could be due to the nature and framing of the questions. When engaging in socially desirable responding, an individual taking the dark triad must under claim their negative traits rather than over claim their positive ones. With the exception of narcissism, none of the time 2 score after the warning message claimed significantly less desirable traits than the mass screening true scores, supporting the notion that there is no over-compensatory effect in response to the warning message.

Hypothesis Testing

Hypothesis 1 stated that self-reported guilt would be higher for those in the warned-faker condition than in any of the other three study cells. This hypothesis was partially supported, in that guilt measured by the PFQ scale was significantly higher for warned fakers over those in the control group ($t(120)=4.72, p<.001$), and significantly higher at $p<.10$ for warned fakers over warned honest ($t(120)=1.83, p=.070$). For the PANAS-X measure of guilt, those in the warned condition had a significantly higher mean than those in the control group ($t(120)=5.91, p<.001$). However, while warned fakers did have a higher mean than warned honest, it was not significant ($t(120)=1.64, p=.103$).

Hypothesis 2 stated that self-reported fear would be significantly higher for those in the warned faker condition than any of the other three study cells. This hypothesis was partially supported. Individuals in the warning condition reported significantly higher PANAS-X fear and Fear of Punishment scores than those in the control group ($t(120)= 3.88, p<.001$; $t(120)=10.31, p<.001$). However, those in the warned faker condition did not report higher levels of fear than those in the warned honest condition ($t(120)= 1.31, p=.19$; $t(120)=.432, p=.67$).

Hypothesis 3 stated that self-reported anger would be significantly higher for individuals in the warned honest condition, than in any of the other study cells. This hypothesis was partially supported. Individuals in the warning condition overall reported significantly higher hostility means on the PANAS-X than those in the control conditions ($t(120)=5.76, p<.001$). However those who were in the warned honest condition did not report significantly higher hostility scores than those in the warned faker condition ($t(120)=-1.02, p=.31$).

Hypothesis 4 stated that self-reported fear would be significantly higher for those in the warned-honest condition than either control condition, but not the warned faker condition. This hypothesis was partially supported. Self-reported fear was higher for the warned honest condition than either control condition ((PANAS-X) $t(120)= 3.88, p<.001$; (FoP) $t(120)=10.31, p<.001$), but it was not significantly lower than those in the warned faker condition ($t(120)=1.31, p=.19$; $t(120)=.432, p=.67$).

Hypothesis 5 stated that self-reported guilt fear and anger would be significantly lower in the control condition than the warning condition. This hypothesis was supported for all emotional measures at $p<.001$ (see table 6).

While not directly hypothesized, a general measure of anxiety was also included in the emotions survey. The results showed that those who received the warning message overall reported significantly higher anxiety than those in the control condition (see table 6). Interestingly, there was also a significant difference in the faker vs. honest control condition. Those who were instructed to reply honestly and received the control message, reported significantly higher anxiety than those who were instructed to fake and received the control message ($t(120)=-2.86, p=.005$).

Table 6: Means and standard deviations of self-reported emotions survey

	Faking Warned		Honest Warned		Faking Control		Honest Control	
	M	SD	M	SD	M	SD	M	SD
State/Trait Anxiety	21.68	4.76	20.65	4.57	12.16	4.50	**14.59	4.92
Guilt PANAS-X	11.19	5.79	9.59	4.94	6.65	1.84	6.98	2.83
Fear PANAS-X	12.44	5.99	11.11	5.25	8.00	3.27	8.08	2.98
Anger PANAS-X	11.34	4.66	12.27	5.34	7.57	2.92	7.66	3.15
Guilt PFQ	†11.39	5.24	9.81	4.31	7.08	2.55	6.90	2.04
Shame PFQ	*22.42	8.75	19.52	7.10	13.62	5.55	13.27	5.51
Fear of Punishment	17.85	4.81	17.44	5.44	8.60	4.47	8.68	3.73

Note. Independent t-tests were performed for warned vs. control and honest vs. faker. All scales were significantly higher for the warned condition vs. the control condition ($p < .001$). Scales that were significantly lower in faker vs. non faker within the separate warned vs. control are designated by † $p < .10$, * $p < .05$, ** $p < .01$. $n = 244$

Expected Physiological Responses

In addition, physiological responses were also examined as additional support to the self-reported responses. The mean difference from baseline to manipulation for each cell can be found in table 7.1 and 7.2. There was an issue with the recording equipment and the skin conductance measure did not record properly. This issue was not identified until the analysis stage. Overall there were very few significant results in regards to the physiological measures.

The expected physiological pattern in correspondence with hypothesis 1 was that for warned fakers who experienced guilt, we should see the following pattern: increased heart rate, increased skin conductance, decrease in PEP, and no change in cardiac output. The physiological results were somewhat inconsistent with what was expected. For warned fakers there was an insignificant increase in heart rate, and a marginally significant increase in PEP ($t(1) = -11$, $p = .058$). There was no significant change in cardiac output. This result should be interpreted with caution though, due to the small sample size.

The expected result in correspondence with hypothesis 2 was that for warned fakers and warned honest participant who experience fear, there should be a decrease in heart rate and stroke volume, and an increase in skin conductance, low frequency, and high frequency heart rate variability responses. For warned fakers, there was an increase in heart rate and a decrease in stroke volume (not significant), and a decrease in LF/VLF and HF/VHF heart rate variability responses. The decrease in LF variability was marginally significant ($t(14) = 2.07$, $p = .058$). For warned honest participants there was a decrease in heart rate, an increase in stroke volume, and a

similar decrease in LF/VLF and HF/VHF heart rate variability responses. None of these patterns were significant.

In concordance with hypothesis 3, the physiological pattern expected for anger, for the warned honest group, was a decrease in heart rate, PEP, and SV, and an increase in skin conductance, CO, and RSA. Consistent, there was a decrease in heart rate and an increase in CO, although not significant. The results for PEP, SV, and RSA were all in the opposite direction of what was expected, but similarly not significant.

The last expected physiological pattern was based upon the control group, in that individuals in those cells should feel either no emotions, or a slight increase in anxiety. Therefore the physiological response change should either be none or an increase in heart rate, skin conductance, and cardiac output, but no change in SV. For control honest there was a significant decrease in LF ($t(10)=3.02$, $p=.013$), but no significant changes for any other measure. For those in the control faker condition, there were marginally significant decreases in LF ($t(13)=2.07$, $p=.059$) and VLF ($t(13)=1.87$, $p=.085$), but no significant changes in the other physiological measures.

Due to the lack of significant and consistent findings in regards to the physiological measures in support of the self-reported emotions, the self-reported measures were further analyzed with more rigorous statistics to gain a better understanding of the underlying processes occurring.

Table 7.1: Mean differences in physiological responses between the baseline and warning message.

	SCR		CO		PEP		SV	
	M	SD	M	SD	M	SD	M	SD
Warned Faker ^a	-	-	-.91	1.29	†.006	.001	-12.44	13.34
Warned Honest ^b	-	-	.42	.99	.010	.011	7.88	15.42
Control Faker ^c	-	-	4.02	10.41	.022	.045	51.96	138.27
Control Honest ^d	-	-	-3.93	5.35	.007	.004	-35.65	54.54

Note. ^an=2. ^bn=4. ^cn=11. ^dn=3. Paired samples t-tests were performed for means from baseline to manipulation for each of the four study cells. †p<.10, *p<.05, **p<.01. The equipment did not record properly for SCRs. A negative value indicates a decrease from baseline to manipulation.

Table 7.2: Mean differences in physiological responses between the baseline and warning message.

	RSA		LF		VLF		HR		HF		VHF	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Warned Faker ^a	-.481	1.76	†-2124.45	3978.51	-476.09	1804.95	1.29	5.22	-602.65	12003.06	-507.62	1402.30
Warned Honest ^b	-.307	1.42	-4280.42	10903.29	-325.14	1091.94	-2.37	4.65	-2929.57	9734.76	-120.57	345.08
Control Faker ^c	-.387	1.49	†-1077.88	1947.61	†-105.52	211.43	2.01	5.35	-1919.43	5140.95	-2393.37	8059.71
Control Honest ^d	-.046	.84	*-974.71	1070.47	-72.82	223.67	-5.51	3.39	-66.28	1166.12	9.59	73.63

Note. ^an=16. ^bn=18. ^cn=18. ^dn=13. Paired samples t-tests were performed for means from baseline to manipulation for each of the four study cells. †p<.10, *p<.05, **p<.01. A negative value indicates a decrease from baseline to manipulation.

Mediation Analyses

The previous results demonstrated the warning message elicited significant emotional reactions as compared to the control group. These exploratory results expand upon the previous hypotheses, in that they test whether these emotional reactions actually mediate the process through which warning messages lead to lower scores at time 2. Mediation was tested using the PROCESS macro in SPSS for all possible models in which x= warning vs. control, m=emotions, and y=personality score at time 2, controlling for time 1 (Hayes, 2013).

Significant mediation results were found for the models found in table 8. The most salient emotional response was Fear of Punishment, which mediated all of the scales measured. Guilt was almost as salient and was found to mediate all scales expect for psychopathy. Interestingly, hostility was only found to mediate the Dark Triad scales.

Table 8: Mediation results for middle message, emotions, and personality scores

Dependent Variable	Mediator		Effect		
			Effect	LLCI	ULCI
Conscientiousness	Fear of Punishment	Indirect	-5.28	[-8.76	-2.56]
		Direct	-3.56	[-7.56	.43]
	State Anxiety	Indirect	-2.60	[-4.99	-.53]
		Direct	-6.25	[-10.04	-2.46]
	Shame (PFQ scale)	Indirect	-1.75	[-3.53	-.25]
		Direct	-7.09	[-10.40	-3.78]
	Guilt (PANAS scale)	Indirect	-1.81	[-3.82	-.41]
		Direct	-7.04	[-10.18	-3.89]
Impression Management	Fear of Punishment	Indirect	-2.62	[-4.46	-.93]
		Direct	-1.06	[-3.22	1.11]
	Guilt (PANAS scale)	Indirect	-.86	[-1.94	-.16]
		Direct	-2.82	[-4.52	-1.13]
Narcissism	Guilt (PANAS scale)	Indirect	-.40	[-.78	-.08]
		Direct	-.47	[-1.21	.27]
	State Anxiety	Indirect	-.52	[-1.07	-.01]

		Direct	-.35	[-1.24	.54]
	Fear of Punishment	Indirect	-.92	[-1.67	-.29]
		Direct	.06	[-.90	1.01]
	Guilt (PFQ)	Indirect	-.39	[-.81	-.02]
		Direct	-.48	[-1.23	.28]
	Shame (PFQ)	Indirect	-.42	[-.86	-.04]
		Direct	-.45	[-1.23	.33]
	Hostility	Indirect	-.39	[-.81	-.05]
		Direct	-.48	[-1.25	-.05]
Psychopathy	Fear of Punishment	Indirect	-.72	[-1.44	-.11]
		Direct	-.58	[-1.47	.31]
	Hostility	Indirect	-.37	[-.96	-.03]
		Direct	-.92	[-1.64	-.20]
Machiavellianism	Fear of Punishment	Indirect	-1.01	[-1.84	-.25]
		Direct	-.74	[-1.82	.33]
	Hostility	Indirect	-.49	[-.97	-.10]
		Direct	-1.27	[-2.12	-.41]
	Guilt (PANAS scale)	Indirect	-.49	[-.94	-.14]
		Direct	-1.27	[-2.10	-.44]

Note: Dependent variables for each scale are personality scores at time 2, controlling for scores at time 1.

Moderated Mediation

The following tables demonstrate significant moderated mediation results that were found. As an exploratory measure, all various combinations of emotions and score reduction were tested for moderated mediation. Other than the general model structure, none of the following results were hypothesized a priori, and should therefore be a basis for future directions, rather than confirmation of any testable hypotheses.

As can be seen in Table 9, the interaction between the middle warning manipulation and the instruction set significantly moderated conscientiousness score reduction, and this process was mediated by state anxiety. The index of moderated mediation was significant (-2.40, SE=.96, 95% CI [-4.71, -.81]). The simple effects analysis suggests that there is a significant positive

effect on time 2 conscientiousness scores (controlling for time 1), when fakers are given the control message and a non-significant negative effect when fakers are given the warning.

This suggests that when individuals are warned, the instructions have no significant effect on score reduction. Both the instructed honest and instructed faking participants had significant score reduction due to the warning. When individuals are assigned the control message, the differential effects of the instruction set become significant. Individuals who are instructed to fake and given the control message show less score reduction (continue faking) because they have less anxiety. Individuals who are asked to be honest and given the control message however, show increased anxiety, which in turn leads to less score reduction. As explained previously, it is believed that this anxiety in honest control individuals stems from having to honestly acknowledge the negative personality traits they possess. This result was consistent across all of the other personality variables, as well as the impression management measure. Those results can be seen in tables 10-13.

Table 9: Moderated mediation for conscientiousness and state anxiety

Variable	Mediator = State Anxiety		Dependent variable = Conscientiousness T2	
	β	SE	β	SE
Constant	17.26**	2.27	18.15**	6.03
Instructions	-2.13*	.91	-3.92*	1.71
Middle Message	6.07**	.85		
Instructions X Middle Message	3.58**	1.20		
Conscientiousness T1 (control)	-.02	.02	.93**	.04
State Anxiety			-.67**	.12
<i>F</i>	45.75**		208.07**	
<i>R</i> ²	.43		.72	
Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI	
Direct effect of Instructions on Conscientiousness T2	-3.92	[-7.29		-.55]

Conditional indirect effects of Instructions on
Conscientiousness T2 at:

Control	1.43	[.33	3.02]
Warning	-.97	[-2.43	.17]
	Index		
Mediator: State anxiety	-2.40	[-4.71	-.81]

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. * $p < .05$, ** $p < .01$

Table 10: Moderated mediation for impression management and state anxiety

Variable	Mediator = State Anxiety		Dependent variable = IM T2	
	β	SE	β	SE
Constant	16.27**	1.69	13.84**	2.52
Instructions	-2.15*	.92	-1.26	.93
Middle Message	6.11**	.85		
Instructions X Middle Message	3.53**	1.20		
IM T1 (control)	-.04	.04	.77**	.05
State Anxiety			-.28**	.07
<i>F</i>	45.59**		115.02**	
<i>R</i> ²	.43		.59	
Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI	
Direct effect of Instructions on IM T2	-1.26	[-3.09	.56]	
Conditional indirect effects of Instructions on IM T2 at:				
Control	.61	[.11	1.41]	
Warning	-.39	[-1.05	.08]	
	Index			
Mediator: State anxiety	-.99	[-2.09	-.33]	

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. IM= impression management. * $p < .05$, ** $p < .01$

Table 11: Moderated mediation for Machiavellianism and state anxiety

Variable	Mediator = State Anxiety		Dependent variable = Machiavellianism T2	
	β	SE	β	SE
Constant	16.09**	1.10	4.34**	.91
Instructions	-2.17*	.88	.42	.42

Middle Message	6.13**	.85		
Instructions X Middle Message	3.51**	1.20		
Machiavellianism T1 (control)	-.10	.06	.81**	.04
State Anxiety			-.12**	.03
<i>F</i>		46.25**		156.01**
<i>R</i> ²		.44		.66

Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI
Direct effect of Instructions on Machiavellianism T2	.42	[-.40	1.25]
Conditional indirect effects of Instructions on Machiavellianism T2 at:			
Control	.26	[.06	.61]
Warning	-.16	[-.44	.02]
		Index	
Mediator: State anxiety	-.42	[-.88	-.14]

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. **p*<.05, ***p*<.01

Table 12: Moderated mediation for narcissism and state anxiety

Variable	Mediator = State Anxiety		Dependent variable = Narcissism T2	
	β	SE	β	SE
Constant	14.35**	.84	2.62**	.65
Instructions	-2.60**	.87	-.23	.36
Middle Message	6.05**	.85		
Instructions X Middle Message	3.57**	1.20		
Narcissism T1 (control)	.03	.06	.87**	.04
State Anxiety			-.09**	.03
<i>F</i>		45.17**		213.17**
<i>R</i> ²		.43		.73

Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI
Direct effect of Instructions on Narcissism T2	-.23	[-.94	.49]
Conditional indirect effects of Instructions on Narcissism T2 at:			
Control	.22	[.05	.50]
Warning	-.08	[-.28	.05]
		Index	
Mediator: State anxiety	-.31	[-.68	-.07]

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. *p<.05, **p<.01

Table 13: Moderated mediation for psychopathy and state anxiety

Variable	Mediator = State Anxiety		Dependent variable = Psychopathy T2	
	β	SE	β	SE
Constant	17.95**	1.63	1.95*	.99
Instructions	-2.41**	.85	-.29	.33
Middle Message	6.07**	.84		
Instructions X Middle Message	3.78**	1.20		
Psychopathy T1 (control)	-.17*	.08	.95**	.04
State Anxiety			-.10**	.03
<i>F</i>		47.25**		180.14**
<i>R</i> ²		.44		.69

Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI
Direct effect of Instructions on Psychopathy T2	-.29	[-.95	.36]
Conditional indirect effects of Instructions on Psychopathy T2 at:			
Control	.24	[.07	.54]
Warning	-.13	[-.39	.01]
	Index		
Mediator: State anxiety	-.37	[-.82	-.13]

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. *p<.05, **p<.01

Table 14: Moderated mediation for psychopathy and PANAS guilt

Variable	Mediator = Guilt (PANAS)		Dependent variable = Psychopathy T2	
	β	SE	β	SE
Constant	9.46**	1.45	1.36	.93
Instructions	-.24	.75	-.12	.33
Middle Message	2.61**	.75		
Instructions X Middle Message	2.09*	1.07		
Psychopathy T1 (control)	-.13	.07	.95**	.04
Guilt (PANAS)			-.13**	.04
<i>F</i>		13.16**		179.18**
<i>R</i> ²		.18		.69

Direct, indirect, and index of moderated mediation	Effect	LLCI	ULCI
Direct effect of Instructions on Psychopathy T2	-.12	[-.78	.53]
Conditional indirect effects of Instructions on Psychopathy T2 at:			
Control	.03	[-.06	.20]
Warning	-.24	[-.75	-.02]
	Index		
Mediator: Guilt (PANAS)	-.27	[-.82	-.02]

Note. Instructions and middle messages were dummy coded (honest=0, faking=1; control=0, warning=1). Confidence intervals were calculated using bootstrapping with 5000 repetitions. LLCI=lower limit confidence interval, ULCI=upper level confidence interval. * $p < .05$, ** $p < .01$

The only other moderated mediation that differed can be found in table 14, in which guilt was found to be the mediator for psychopathy. The index of moderated mediation was significant (-.27, SE=.19, 95% CI [-.82, -.02]) and the simple effects analysis demonstrate that similarly, for those who were given faking instructions, the effect of the control message was positive on time 2 scores and negative on time 2 scores for those who were warned. What is different about this result, however, is that the effect of the warning message, and not the control message is significant and this process is mediated through guilt and not anxiety. What this result suggests is that when individuals are warned, the instruction set has a significant effect on psychopathy score reduction through guilt. When individuals are given the control message, the instruction set has no influence on their level of guilt. However when individuals are given the warning message, those who instructed to fake have higher levels of score reduction mediated through their guilt.

Qualitative Results

Due to the concern that the instructed faking being met with a warning message would cause confusion, a short answer question was added in to the end of the emotion reactions

questionnaire. The question asked participants to please add in any emotions that they felt in reaction to the middle message that were not adequately captured in the previous survey. This measure was only included to clarify and give insight into any unusual results that may have occurred.

Those in the honest control condition had no significant emotional reactions as expected, although some simply stated they were confused if something had gone wrong. Some example responses were:

“I did not think much of it, it was kind of just an awkward minute of sitting in a room by myself”

“I was a little confused, but not scared I had failed the test or had done anything wrong.”

For those in the honest warned condition, almost all responded with anger and irritation, as expected. Some however, felt embarrassed at being identified in that manner even though they felt they were being honest. Some example responses were:

“I was irritated because I had answered them as well as I could and felt like I was being called a liar.”

“It made me self-reflect and try to be more introspective in my answering. I also thought it was somewhat embarrassing that I had lied without thinking about lying.”

For those in the faking control condition, the emotional responses were as expected in that individuals either had no reaction, or were slightly anxious at the interruption. Some example responses were:

“I didn't think much of it.”

“I didn't think the message was due to my responses. I felt anxious because of the pause yet did not blame myself for the pause.”

Finally, for those in the faking warned condition, individuals unfortunately did respond with anger or confusion at being told to respond as an ideal employee and then warned against it. Some individuals, however, felt anger because they felt they were being honest, which was contrary to their instruction set. Therefore, the level of attention paid to specific instructions might have been a confounding factor in this study. Some example responses were:

“I was confused because I thought the previous instructions told me to be desirable. So after that I took it truthfully applying to myself.”

“The message made me nervous and angry because I felt that I received it unjustly considering I had answered the questions honestly. I did the survey again in the same manner as the first time for the most part because I knew I had told the truth the first time.”

However, this unanticipated result does support the hypothesis that anger is a primary emotional reaction to being unjustly identified as a faker.

Discussion

Manipulation and over compensatory check

Before examining the main purpose of this study, which was to determine the underlying emotional mechanism behind score reduction, it first needed to be established that the experimental design functioned as intended. The manipulation check results demonstrate that those who were given the faking instructions did report higher levels of impression management than those in the honest condition. Similarly, there was a significant difference between time 1 impression management and the true personality mass screening scores for those who received faking instructions, but not for those who received the honest instructions. This suggests that the different instruction sets were a successful manipulation in producing honest and faking respondents.

The next manipulation check was to ensure that individuals reduced their personality scores after receiving the warning message, but not over compensate and report scores lower than their true personality scores. The results demonstrate that warned individuals did significantly reduce their scores from time 1 to time 2 regardless of instruction set indicating that the warning message is successful in reducing personality scores. Narcissism in the honest warned condition was the only exception to this, in that desirable responding was reduced but not significantly in time 2. This could be because narcissism is not necessarily seen as a bad trait to have, particularly in the organizational setting, and therefore was not being downplayed as an undesirable trait.

A similarly unexpected result was that fakers who received the control message significantly reduced their scores on conscientiousness and impression management. This could be due to a number of reasons such as a higher item count, which allows for more variability in responses on those measures, and therefore there is less range restriction in reducing those scores.

Comparing the scores of time 1, time 2, and the individual's true personality scores show that even when individuals are given the honest instruction set, their personality scores at time 1 are significantly higher than their true scores for all measures except impression management; a direct measure of faking. Therefore, it seems that individuals when put into a salient testing context, such as in a lab setting, they fake regardless of instruction set. It is then no surprise that there was score reduction after the warning message for both instruction sets.

When comparing participant's true personality scores to their time 2 scores, after the warning or control message, results show that for the control message there was significantly more desirable responses for all scales except narcissism. For the warning message, results show that there was no significant difference between the individuals' true personality score and their time 2 scores on impression management, Machiavellianism, and psychopathy. This supports that there is no over compensatory response for these scales.

However, participants reported significantly higher conscientiousness scores than their true personality score, but significantly lower narcissism scores than their true personality. The high conscientiousness score could be due to the context in which they were filling the personality surveys out. Because the conscientiousness survey is school specific, taking the test in a lab setting at school could make their responses more salient than taking the test at home,

outside of a school setting. However, overall this result also supports the notion that there is no over compensatory effect.

As for narcissism, there is a significant decrease in socially desirable responses for those who were warned, meaning that individuals were claiming more narcissistic tendencies in time 2 than their true personality score. This result suggests that there could be a slight over compensatory effect for narcissism, or an underreporting of narcissism in the mass screening baseline score.

Emotional Reaction to Warning Message

As expected the warning message elicited significantly higher negative emotional reactions from participants than the control message for all emotion measures. When differentiated between honest and faking instruction sets, guilt was shown to be marginally higher for warned fakers than for warned non-fakers. As was expected, those who were asked to be honest and given the warning message reported the highest levels of anger, although it was not significantly different from those who were asked to fake and given the warning. However, one emotional measure that was not hypothesized, but was included in the study found that those in the warned faker condition reported significantly higher shame than those who were given honest instructions and warned.

One results found that was unexpected, but interesting, was that those who were given honest instructions and shown the control message reported significantly higher levels of anxiety than those who were told to fake and given the control message. As the control message is a neutral stimulus, there should not be any emotional differences between these two groups. I believe that the reason this occurred was due to the nature of the personality test. The dark triad

measures are negative qualities that are not considered socially desirable to claim. I believe that asking individuals to respond as honestly as possible, and therefore, to claim these very negative traits, increased their state anxiety. Those who were asked to fake did not have to claim these negative qualities. I believe that this result further supports that participant's response behaviors differ depending on the content and framing of the personality measure.

Physiological Support

Unfortunately the physiological recording equipment did not cooperate as was hoped. The recording of impedance cardiography would often cut in and out during testing, and would have required frequent interruptions of the test-takers progress in order to remedy. Although this is the recording that many of the expected measures were based on, it was decided that it should be left alone in order to not distract and disrupt the testing process, and thereby influence the other physiological measures. Therefore, the final sample sizes for many of the measures were not adequate to properly interpret.

For those that were, there was only one significant result at the $p < .05$ level, which was a decrease in low frequency heart rate variability for honest control. This was contrary to any of the expected result patterns. Therefore, the conclusions drawn will be based solely on the self-reported emotions. Due to this lack of physiological support, more complicated statistics that were not previously hypothesized were performed, in order to gain insight into the processes occurring.

Mediations

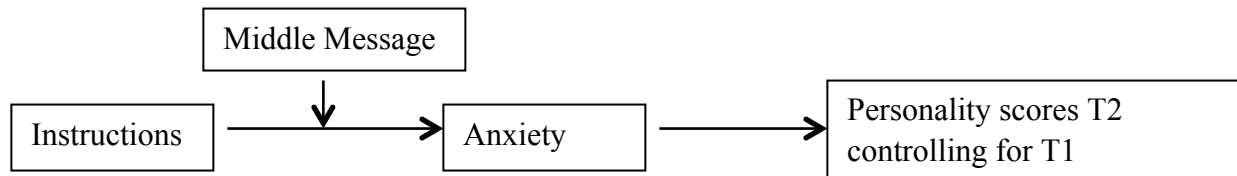
The most prominent result found was that fear of punishment significantly mediated all of the personality scales. This suggests this emotional reaction is the primary mechanism behind

the reduction in socially desirable responding in response to a warning message. Guilt was found to mediate this score reduction in every scale except psychopathy, suggesting it is also a salient emotion in response to the warning message. One really interesting result that stood out from the mediation analyses was that hostility was found to mediate the decrease in socially desirable responding for only the Dark Triad measures. This result once again suggests that individuals might have different behavioral, cognitive, and emotional responses to the dark triad measures of personality as compared to more traditional measures.

Moderated Mediations

Moderated mediation was found for all personality scales in which the warning vs. control manipulation interacted with the instruction set to result in reduced socially desirable responding, mediated by anxiety. The model can be seen in figure 1.

Figure 1. Moderated mediation model



The results of the moderated mediation results stem back to the higher reported anxiety in the honest control condition that was previously reported and is the same for all personality scales, and therefore will be interpreted all together. For all personality scales, the instruction set had a negative coefficient with state anxiety, meaning that asking individuals to fake reduced the level of anxiety, as opposed to asking them to respond honestly. Within the moderated mediation, the warning message (moderator) had a positive coefficient with state anxiety, suggesting that the

warning increases levels of anxiety. In step 2 of the moderated mediation, state anxiety had a significant negative coefficient with personality at time 2, suggesting that a higher level of anxiety will produce a lower personality score at time 2. When we look at the conditional indirect effects of instructions on personality time 2 scores, we see that of the moderator, only the control message had a significant effect. This result demonstrates that there is a positive effect on T2 conscientiousness scores when individuals are told to fake and then given the warning message. This effect is mediated by state anxiety. Essentially control fakers have less reported anxiety, and therefore do not have that negative effect on time 2 personality scores as a result from anxiety. The effect of the warning message on fakers had a negative effect on time 2 personality scores through anxiety, but this indirect effect was not significant.

A similar result was found in which guilt mediated the moderation of instructions and middle message on psychopathy score reduction. For those who were given faking instructions, the effect of the control message was positive on time 2 scores and negative on time 2 scores for those who were warned. As opposed to the anxiety results, the simple effect of the warning message, and not the control message was significant. This suggests that the effect of the warning message had a stronger significant effect on fakers, and their score reduction on psychopathy was mediated through guilt.

Contributions

The present study makes several contributions to understanding the emotional mechanisms behind the reduction in socially desirable responding in reaction to a middle warning message. First it identifies fear of punishment as a primary emotional reaction and mechanism for score reduction. The warning message produced more honest personality

responses, and this process was mediated by fear of punishment for all four personality measures and the impression management measure.

Second, the differential results between conscientiousness and the Dark Triad measures suggest that individuals' test taking behavior and response styles differ between the more traditional Big 5 measures and this newer negative personality test. Hostility was found as a mediator between the warning message and score reduction for all of the Dark Triad measures, but not for conscientiousness or impression management. This suggests that the emotional reactions in response to these tests differ and more research should be done to identify if fairness reactions and face validity differ for Dark Triad measures.

Third, this study included a baseline measure of personality in order to examine if individuals were underreporting their personality scores out of fear after being given the warning message. This study demonstrated that with the possible exception of narcissism, individuals did not significantly underreport their scores, and in the case of conscientiousness, continued to report high levels of the positive trait even after being warned.

Finally this study was the first to examine state anxiety as a possible emotional mechanism. This measure was found to play a significant role in the emotional and behavioral processes. It was a significant mediator in the moderated mediation models and was found to have important implications in explaining the emotional processing for those who were asked to be honest in reporting their dark personality traits.

Limitations and Future Research

One of the primary limitations of this study was the instructed faking set. When the study was first being designed, an instruction faking set seemed as though it would be a benefit in that

it would make emotional reactions such as anger more salient, it had been used in previously published research, it would even the study cells, and it was a less arbitrary way to classify fakers than using a faking scale. However, this influenced the emotional reactions of those in the warned faker cell, as some individuals met the warning message with confusion.

A qualitative measure was included to have individuals expand on their emotional reactions to the middle message, and as anxiously anticipated, some individuals in the warned faker condition did report confusion and irritation. However, there were also those in the same group that reported anger because they felt they were responding honestly. This suggests that perhaps the instruction set was not headed as strongly as hoped. One potential way to get around this issue could be asking individuals to fake in a way that they won't get caught. Future research with similar study design should weigh the pros and cons of utilizing a deception component and faking measures vs. the instructed faking set.

Another limitation was that the physiological recording equipment did not record as perfectly as expected. Many of the differentiating measures are based upon one type of recording equipment, so future researchers should determine the best possible setup to insure an uninterrupted recording, or have a plan in place for adjusting electrodes if the recording cuts out.

Lastly, as this study was conducted with a student sample in a lab setting, in which they were attached to physiological recording equipment, it is difficult to say that the results will generalize to typical applicants. Most applicants are motivated to fake and taking their personality tests either at home or in a testing center, and therefore, do not have the same pressures or lack of comfort.

References

- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12), 1031-1051.
- Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, 69, 574-580.
- Barrett, K. C. (1995). A functionalist approach to shame and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. (pp. 25–63). New York, NY: Guilford.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81, 261-272.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and Assessment*, 9, 9-30.

- Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, *43*(1), 93-103.
- Berntson, G. G., Quigley, K. S., & Lozano, D. (2007). Cardiovascular psychophysiology. *Handbook of psychophysiology*, *3*, 182-210.
- Bing, M. N., Kluemper, D., Davison, H. H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, *116*, 148-162.
- Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternative Predictors, and Job Performance. *Personnel Psychology*, *52*, 561-589.
- Bradley, M. M., & Lang, P. J. (2000). Emotion and motivation. *Handbook of Psychophysiology*, *2*, 602-642.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49-59.
- Burns, G. N., Fillipowski, J. N., Morris, M. B., & Shoda, E. A. (2015). Impact of electronic warnings on online personality scores and test-taker reactions in an applicant simulation. *Computers in Human Behavior*, *48*, 163-172.
- Butcher, J. N., Morfitt, R. C., Rouse, S. V., & Holden, R. R. (1997). Reducing MMPI-2 defensiveness: The effect of specialized instructions on retest validity in a job applicant sample. *Journal of Personality Assessment*, *68*, 385-401.

- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of Emotions*, 2, 173-191.
- Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychological bulletin*, 135(2), 183.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18, 267-307.
- Codispoti, M., & De Cesare, A. (2007). Arousal and attention: Picture size and emotional reactions. *Psychophysiology*, 44(5), 680-686.
- Dimberg, U. (1986). Facial reactions to fear-relevant and fear-irrelevant stimuli. *Biological psychology*, 23(2), 153-161.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level test taker faking using the randomized response technique. *Human Performance*, 16, 81-106.
- Dullaghan, T. R. (2010). The effect of a reasoning warning on faking in personality testing for selection and the perception of procedural justice (Doctoral dissertation, University of South Florida).
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169-200.
- Elaad, E. (2009). Effects of context and state of guilt on the detection of concealed crime information. *International Journal of Psychophysiology*, 71(3), 225-234.
- Ellingson, J. E., Heggstad, E. D., & Makarius, E. E. (2012). Personality retesting for managing intentional distortion. *Journal of Personality and Social Psychology*, 102, 1063.

- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of test taker comparison and construct validity. *Journal of Applied Psychology*, 84, 155.
- Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, 97, 866.
- Fan, J., Zhao, G., Wang, M., Page, R. C., Meng, H., Yuan, Y. (2015). A Comprehensive Examination of a New Faking-Mitigation Procedure for Personality Tests within Selection Contexts: A Field Experiment. SIOP Small Grant Proposal.
- Fourie, M. M., Rauch, H. G., Morgan, B. E., Ellis, G. F., Jordaan, E. R., & Thomas, K. G. (2011). Guilt and pride are heartfelt, but not equally so. *Psychophysiology*, 48(7), 888-899.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, 7(3), 199-216.
- Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of psychophysiology*, 3(1), 51-64.
- Hack, E. M., Yao, X., Page, R. C., Yuan, Y., & Fan, J. (2015, April). SDE and IM scores moderate personality validities among job applicants. Paper to be presented as part of the symposium conducted by Fan, J. (Chair), Investigating validities of several faking measures within selection contexts at the 29th annual conference of the Society of Industrial and Organizational Psychology, Philadelphia, PA.

- Handler, M., & Honts, C. R. (2007). Psychophysiological mechanisms in deception detection: A theoretical overview. *Polygraph*, 36(4), 221-232.
- Harder, D. H., & Zalma, A. (1990). Two promising shame and guilt scales: A construct validity comparison. *Journal of Personality Assessment*, 55, 729–745.
- Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression based approach. New York, NY: Guilford Press.
- Heaphy, E. D., & Dutton, J. E. (2008). Positive social interactions and the human body at work: Linking organizations and physiology. *Academy of Management Review*, 33(1), 137-162.
- Height/Weight entries for CO using Kubicek/Rho method | BIOPAC. (n.d.). Retrieved April 12, 2018, from <https://www.biopac.com/knowledge-base/heightweight-entries-for-co-using-kubicekrho-method/>
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, 1, 272-290.
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological assessment*, 22(2), 420.
- Kemper, T. D. (1987). How many emotions are there? Wedding the social and the autonomic components. *American Journal of Sociology*, 263-289.
- Khalifa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, 328 (2), 145-149.

- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84, 394-421.
- 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.
- Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, 6(2), 100-112.
- Lewis, M. (2000). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions*. (pp. 623–636). New York, NY: Guilford Press.
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(3), 301-306.
- O'Boyle Jr, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the dark triad and work behavior: A social exchange perspective. *Journal of Applied Psychology*, 97(3), 557.
- Overbeek, T. J., van Boxtel, A., & Westerink, J. H. (2012). Respiratory sinus arrhythmia responses to induced emotional states: effects of RSA indices, emotion induction method, age, and sex. *Biological Psychology*, 91(1), 128-141.
- Pace, V. L., & Borman, W. C. (2006). The use of warning to discourage faking on noncognitive inventories. A closer examination of test taker faking behavior, 283-304.

- Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package*, 41.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. *The role of constructs in psychological and educational measurement*, 4969.
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, 84, 890-904.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of research in personality*, 36(6), 556-563.
- Pinsky, M. R. (2003). Rationale for cardiovascular monitoring. *Current opinion in critical care*, 9(3), 222-224.
- Rauthmann, J. F. (2011). Acquisitive or protective self-presentation of dark personalities? Associations among the Dark Triad and self-monitoring. *Personality and Individual Differences*, 51(4), 502-508.
- Reyes del Paso, G. A., Langewitz, W., Mulder, L. J., Roon, A., & Duschek, S. (2013). The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies. *Psychophysiology*, 50(5), 477-487.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80(5), 607.

- Stemmler, G., Heldmann, M., Pauls, C. A., & Scherer, T. (2001). Constraints for emotion specificity in fear and anger: The context counts. *Psychophysiology*, *38*(2), 275-291.
- Tangney, J. P. (1990). Assessing individual differences in proneness to shame and guilt: development of the Self-Conscious Affect and Attribution Inventory. *Journal of Personality and Social Psychology*, *59*, 102.
- Tangney, J. P. (1999). The self-conscious emotions: Shame, guilt, embarrassment, and pride. In T. Dalgliesh & M. J. Power (Eds.), *Handbook of Cognition and Emotion*. (pp. 541–568). Chichester, England: Wiley.
- van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*, 197-210.
- Watson, J. (1924). *Behaviorism*. New York: People's Institute.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*.

Appendix A Measures for Current Study

School-specific Conscientiousness scale

Instruction: Please read the following statements and indicate the degree to which each statement is an accurate description of you with respect to how you behave. Describe yourself as you *generally* are (most of the time, in most situations), not as you wish to behave in the future. Describe yourself as you honestly see yourself in relation to others who are of the same sex as you, and roughly your same age. Your responses will be kept in absolute confidential, so please respond honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral/ Undecided	Slightly Agree	Agree	Strongly Agree

1. I'm pretty good about pacing myself so as to get assignments done on time.
2. I don't seem to be completely successful at anything school-related.
3. I'm something of a "workaholic" at school.
4. When I start a self-improvement program to improve my study habits, I usually let it slide after a few days.
5. Once I start a class project, I almost always finish it.
6. When it comes to class participation and course work, I am easy-going and lackadaisical.
7. When a term project gets too difficult, I'm inclined to start a new one.
8. I strive for excellence in everything I do at school.
9. I'm known for my prudence and common sense when it comes to academic activities.
10. When I study, I waste a lot of time before settling down to work.
11. I often come into academic situations (e.g., tests, group projects, etc.) without being fully prepared.
12. I am efficient and effective at my schoolwork.
13. I have a lot of self-discipline in my course work.
14. In my college studies, I don't feel like I'm driven to get ahead.
15. I'm a very competent person when it comes to college level work.
16. There are so many little class assignments that need to be done that I sometimes just ignore them all.
17. For school projects, I have a clear set of goals and work toward them in an orderly fashion.
18. I don't take student duties like completing professor evaluations very seriously.
19. I strive to achieve all I can at college.
20. In college, I am a productive person who always gets the work done.

21. I keep myself informed and usually make intelligent decisions on class-related projects (e.g., long-range assignments, term papers, group projects, etc.).
22. I have trouble making myself do what I should for classes.
23. I pride myself on my sound judgment in my college class work.
24. I work hard to accomplish my academic goals (e.g., completing homework, making grades, etc).

Impression Management Scale

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral/ Undecided	Slightly Agree	Agree	Strongly Agree

Reverse
Scored

1. I am sometimes irritated by people who ask favors of me. X
2. I sometimes try to get even, rather than forgive and forget. X
3. I am always courteous, even to people who are disagreeable.
4. Sometimes at elections I vote for candidates I know little about. X
5. When I take sick-leave from work or school, I am always as sick as I say I am.
6. I always apologize to others for my mistakes.
7. Once in a while I laugh at a dirty joke. X
8. I always tell the truth.
9. There have been occasions when I have taken advantage of someone. X
10. I have never dropped litter on the street.

Dark Triad

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral/ Undecided	Slightly Agree	Agree	Strongly Agree

Subscale

1. I tend to manipulate others to get my way. M
2. I have used deceit or lied to get my way. M
3. I have use flattery to get my way. M
4. I tend to exploit others towards my own end. M
5. I tend to lack remorse. P
6. I tend to be unconcerned with the morality of my actions. P
7. I tend to be callous or insensitive. P
8. I tend to be cynical. P
9. I tend to want others to admire me. N
10. I tend to want others to pay attention to me. N
11. I tend to seek prestige or status. N
12. I tend to expect special favors from others. N

PFQ2 Guilt and Shame Scale Harder (1990)

Embarrassed	G
Mild guilt	S
Feeling ridiculous	S
Worry about hurting or injuring someone	G
Self-consciousness	S
Felling humiliated	S
Intense guilt	G
Feeling "stupid"	S
Regret	G
Feeling "childish"	S
Feeling helpless, paralyzed	S
Feelings of blushing	S
Feeling you deserve criticism for what you did	G
Feeling laughable	S
Feeling disgusting to others	S
Remorse	G

PANAS-X

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you felt this way during the test interruption. Use the following scale to record your answers:

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

- | | | | |
|--|---------------------------------|----------------------------------|---|
| <input type="checkbox"/> cheerful | <input type="checkbox"/> sad | <input type="checkbox"/> active | <input type="checkbox"/> angry at self |
| <input type="checkbox"/> disgusted | <input type="checkbox"/> calm | <input type="checkbox"/> guilty | <input type="checkbox"/> enthusiastic |
| <input type="checkbox"/> attentive | <input type="checkbox"/> afraid | <input type="checkbox"/> joyful | <input type="checkbox"/> downhearted |
| <input type="checkbox"/> bashful | <input type="checkbox"/> tired | <input type="checkbox"/> nervous | <input type="checkbox"/> sheepish |
| <input type="checkbox"/> sluggish | <input type="checkbox"/> amazed | <input type="checkbox"/> lonely | <input type="checkbox"/> distressed |
| <input type="checkbox"/> daring | <input type="checkbox"/> shaky | <input type="checkbox"/> sleepy | <input type="checkbox"/> blameworthy |
| <input type="checkbox"/> surprised | <input type="checkbox"/> happy | <input type="checkbox"/> excited | <input type="checkbox"/> determined |
| <input type="checkbox"/> strong | <input type="checkbox"/> timid | <input type="checkbox"/> hostile | <input type="checkbox"/> frightened |
| <input type="checkbox"/> scornful | <input type="checkbox"/> alone | <input type="checkbox"/> proud | <input type="checkbox"/> astonished |
| <input type="checkbox"/> relaxed | <input type="checkbox"/> alert | <input type="checkbox"/> jittery | <input type="checkbox"/> interested |
| <input type="checkbox"/> irritable | <input type="checkbox"/> upset | <input type="checkbox"/> lively | <input type="checkbox"/> loathing |
| <input type="checkbox"/> delighted | <input type="checkbox"/> angry | <input type="checkbox"/> ashamed | <input type="checkbox"/> confident |
| <input type="checkbox"/> inspired | <input type="checkbox"/> bold | <input type="checkbox"/> at ease | <input type="checkbox"/> energetic |
| <input type="checkbox"/> fearless | <input type="checkbox"/> blue | <input type="checkbox"/> scared | <input type="checkbox"/> concentrating |
| <input type="checkbox"/> disgusted with self | <input type="checkbox"/> shy | <input type="checkbox"/> drowsy | <input type="checkbox"/> dissatisfied with self |

State-Trait Anxiety Inventory

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

1. I feel calm
2. I am tense
3. I feel upset
4. I am relaxed
5. I feel content
6. I am worried

Fear of Punishment Scale

Instructions: The following statements describe how you felt during the psychological test. Please indicate the extent to which you agree or disagree with each statement using the following scale.

1	2	3	4	5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

1. The system message I received in the middle of the test made me very concerned about possibly failing the psychological test.
2. After receiving the system message in the middle of the test, I felt I had to change my response patterns in order pass the psychological test.
3. I felt that if I had continued my response patterns after receiving the system message in the middle of the test, I would have failed the psychological test.
4. The system message I received in the middle of the test did not affect my subsequent response pattern during the rest of the psychological test. (R)
5. The system message I received in the middle of the test did not worry me. (R)

Appendix B Debriefing Statement

Thanks very much for your participation in this study. The aim of this study is to examine faking under an application condition. As you know, applicants tend to enhance themselves and respond in a more socially desirable way when applying for a job, a phenomenon called “social desirability” or “faking.” There is some evidence that faking may render the selection questionnaires invalid. Therefore you were randomly assigned either to complete the questionnaire honestly, or in order to make yourself look like an ideal applicant. Participants were then randomly assigned to receive a control message or a warning message indicating that faking was detected. We will compare the responses of the warned subjects against the unwarned subjects or the control group. If you received the fake-good instructions, you were given a slight deception in order to motivate you to perform well. You were told you would receive \$25 for being in the top pool of “applicants”. There will still be a chance for you to receive this \$25, however it will be out of a random cash drawing in order to ensure fairness to those who received the honest condition, which had no such motivation component. We will do the drawing by the end of this semester. Each participant will get an equal chance to win the cash rewards.

Given that this study involves a small deception, you have the choice of not allowing us to use your data in further analysis. If you choose this option, please notify the experimenter now. In this case, you will still receive 1 hour of SONA credit and be entered into the cash drawing. If you feel upset by the deception, you are more than welcome to have a follow-up with either Dr. Fan or a counselor.

Lastly we ask that you do not inform your peers of this small deception, as they may be future participants and we hope that they will similarly use it as motivation to perform well.