ABSTRACT

Title of Thesis: WARNING AGAINST FAKING: BOON OR BANE, EFFECTS OF WARNING ON TEST SCORE VALIDITY

Junjie Su, Master of Science, 2016

Thesis Directed By: Professor Paul Hanges

Department of Psychology

Previous research found personality test scores to be inflated on average among individuals who were motivated to present themselves in a desirable fashion in high stakes situations, such as during the employee selection process. One apparently effective way to reduce the undesirable test score inflation in such situations was to warn participants against faking. This research set out to investigate whether warning against faking would indeed affect personality test scores in the theoretically expected fashion. Contrary to expectations, the results did not support the hypothesized causal chain. Results across three studies show that while a warning may lower test scores in participants motivated to respond desirably (i.e., to fake), the effect of warning on test scores was not fully mediated by: a reduction in motivation to do well and self-reports of exaggerated responses in the personality test. Theoretical and practical implications are discussed.

WARNING AGAINST FAKING:
BOON OR BANE, EFFECTS OF WARNING ON TEST SCORE VALIDITY

By

Junjie Su

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2016

Advisory Committee:
Professor Paul Hanges, Chair
Professor Michele Gelfand
Associate Professor Edward Lemay

Table of Contents

Chapter 1: Introduction

Despite the prevalence of personality tests as predictors in Industrial-Organizational (I-O) psychology (Barrick & Mount, 1991; Hough & Oswald, 2008), there remains disagreement over whether the usefulness of personality test scores as employee selection tools suffers as a result of distortion on personality tests by job applicants affects (Griffith & Peterson, 2006; Hogan, Barrett, & Hogan, 2007; Morgeson et al., 2007; Ziegler, Maccann, & Roberts, 2012). Given the frequency of applicant faking (Griffith, Chmielowski, & Yoshita, 2007), this thesis examined the extent to which a warning, which is expected to reduce faking on personality tests, might in fact contribute to systematic but irrelevant differences in the personality test responses. Specifically, this thesis attempted to extend theory on applicant faking (Griffith and Peterson, 2011), by emphasizing the need to question the assumption of what is actually defined as faking. More specifically, by using warning as an example of an intervention against faking, I demonstrated that the effectiveness of a given intervention against faking can vary according to how faking is defined.

Initial research on faking, which took the form of multiple lab studies, consistently showed that when instructed to “fake good”, participants can alter their responses. For instance, compared to those who were instructed to respond honestly, those instructed to “fake good” tended to have higher scores (Adair, 2014; Viswesvaran & Ones, 1999). Further research with actual job applicants in the field went on to show that job applicants do score higher on desirable traits in personality tests compared to non-applicants (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). This

accumulation of evidence naturally led researchers to treat higher mean scores on desirable personality traits as an indicator for the presence of faking.

Despite concerns about faking in personality tests, employers continue to include personality tests in employee selection. There are at least three reasons why this could be the case. Firstly, accumulated evidence shows a clear albeit moderate association between personality and various organizationally relevant criteria. For example, personality has been shown to predict diminished counter-productive work behavior (Salgado, 2002) as well as work performance, both at the individual (Barrick, Mount, & Judge, 2001; Dudley, Orvis, Lebiecki, & Cortina, 2006; Judge, Rodell, Klinger, Simon, & Crawford, 2013) and team level (Peeters, Van Tuijl, Rutte, & Reymen, 2006). Secondly, personality tests can also be used in concert with other predictors such as cognitive ability tests, specifically as way to potentially mitigate sub-group differences and reduce adverse impact (Ployhart & Holtz, 2008). Thirdly, it could be also due to the belief that faking can be managed with interventions. Indeed, much effort has been expended to investigate the effectiveness of various interventions against faking.

Considering the variety of interventions against faking, one can make sense of them by sorting them according to their function. This suggests that two types of interventions. There are interventions that try to prevent faking from occurring in the first place, as well as interventions that seek to deal with the effects of faking after it has occurred. While some interventions such as retesting (Ellingson, Heggestad, & Makarius, 2012; Fan et al., 2012; Landers, Sackett, & Tuzinski, 2011), may initially appear belong to neither, because retesting ultimately relies on accurate identification of faking/fakers

before asking them to retest, I would argue that retesting should be considered a type of remedial intervention.

More importantly, not all interventions are equal in terms of effectiveness against faking. Interventions that have been deemed less effective include “correcting” test scores for faking (Christiansen, Goffin, Johnston, & Rothstein, 1994), and removing cases that have been “flagged” (where significant faking is suspected to be present) from consideration altogether (Schmitt & Oswald, 2006). The effectiveness of these remedial interventions has been discounted based on evidence from both simulation (Schmitt & Oswald, 2006) and field data (Christiansen et al., 1994; Hough, 1998). Specifically, neither “correcting” for faking, nor removing individuals that have “flagged” as faking from consideration, had any impact on the ability of the “remediated” personality scores to predict a desired criterion. More crucially, remediation runs the inevitable risk of false positives. Until the potential risk of incorrectly identifying the presence of faking when it is in fact truly absent, can be adequately managed, the justifiability any individual hiring decisions that are based on “adjusted” test scores will continue to remain debatable. Put simply, the observation that “the ethics of denying someone a job opportunity as a result of a ‘correction’ that has no empirical evidence supporting its use is questionable (Goffin & Christiansen, 2003, pg. 343)”, still rings true today.

Not many other interventions have fared any better. For example, researchers have also considered the use of alternative response formats such as such as forced-choice (Jackson, Wroblewski, & Ashton, 2000), in place of multiple-choice, as a way to prevent faking. Unfortunately, subsequent research claimed that the forced-choice format did not protect against faking (Heggestad, Morrison, Reeve, & McCloy, 2006) and may

have harmed test scores. In particular, evidence has suggested that performance on personality test scores using the forced choice format depended on participants' cognitive ability as well (Christiansen, Burns, & Montgomery, 2005; Vasilopoulos, Cucina, Dynomina, Morewitz, & Reilly, 2006). As such, researchers have been forced to develop and explore other interventions.

The sole intervention that has remained relatively unscathed is the use of warning against faking (Burns, Fillipowski, Morris, & Shoda, 2015; McFarland, 2003; Robson, Jones, & Abraham, 2007). To be clear, the claim that warnings are effective against faking was built on data showing lower test scores in the warned compared to no-warning condition. To put it succinctly, because past research indicated that faking led to test score inflation, the effectiveness of an intervention against faking (e.g., warning) was naturally indexed as the opposite effect, i.e., the magnitude of test score deflation.

Indeed, in a meta-analysis of various interventions against faking, the data was consistent with the expectation of lower scores in the warning versus control condition (Adair, 2014; see also Dwight & Donovan, 2003). As mentioned previously, the evolution of faking research has largely led researchers to treat the standardized mean difference (SMD) in scores between the two response instruction conditions: (1) respond honestly and (2) "fake-good" and implicitly treated the SMD as a proxy for faking. As a result, a consistently smaller meta-analytic SMD in the warning compared to the control condition across different personality traits has led to the conclusion that warnings were effective against faking.

Upon initial inspection, the meta-analytic evidence (Adair, 2014; Dwight & Donovan, 2003) seems to indicate that warnings reduced faking, and by extension

increased test score accuracy. However, I propose that it would be premature to conclude that warnings actually reduce faking and/or enhances the accuracy of personality test scores. Put simply, I am arguing that the field needs to move beyond the research designs that it has employed so far. To be more precise, the conclusions that stem from past research have implicitly assumed that the effect of warnings on test scores were *always* due to a reduction in faking. To be clear, while this assumption can be true, it need not always be true. Restated, a warning could hypothetically reduce faking (or not) and/or cause additional effects other than reduced faking as well. To put it bluntly, if one could demonstrate a counter-example that falsifies the claim that lower test scores are always due to reduced faking, then one would have to rule out the claim that the effect of warning is solely limited to reduced faking only.

Plainly speaking, it is vital to recognize that while faking can result in a shift to a higher mean personality score, faking can also manifest in other ways such as variations in (patterns of) item responses (Kuncel & Borneman, 2007; Zickar, Gibby, & Robie, 2004; Ziegler, Maaß, Griffith, & Gammon, in press). In other words, "because different indices inherently make different assumptions about the nature of faking behavior" (Burns & Christiansen, 2011, pg. 368), we should acknowledge the fact that mean differences in test scores is but only one of the ways by which faking may be observed.

As such, I propose that the practice of inspecting if there are mean score differences as a result of warning does not provide sufficient evidence to evaluate whether a warning was effective against faking. For example, there is no reason to rule out the possibility that a warning could reduce faking without affecting test scores. At the same time, we need to consider the possibility that an ineffective (harmful) warning could

"superficially" lower test scores, without achieving the actual desired effect of reducing faking. That is to say, instead of only focusing on whether a warning has effects on test scores, we should broaden our view to examine whether the hypothesized process of warning (reduced faking) actually occurs in practice.

To summarize, whether a warning lowers test score is at best of secondary importance. Of primary importance is the demonstration that a warning actually reduces faking. At the same time, demonstrating that a warning reduces faking is not enough. We need to also rule out the possibility that a warning causes other irrelevant effects apart from a decrease in faking. I submit that only by meeting these two goals would we be justified in claiming that the use of a warning has contributed to an overall increase in test score validity.

The Evidence (So Far)

To simplify, I have labeled the past research designs that has been used thus far to study the effects of warning on faking, the "prevailing standard". Based on a recent meta-analysis of various interventions against faking (Adair, 2014), there are two research designs that have been frequently used to demonstrate that warning is an effective intervention against faking. The first research design uses a one-way ANOVA where participants are randomly assigned to either a warning or a control (i.e., no warning) condition. This design has been used in contexts, such as high stake selection settings, where job applicants are assumed to be motivated to fake (for e.g., Vasilopoulos, Cucina, & McElreath, 2005, Study 1). In this type of research design, a significant main effect of the warning manipulation in which the personality test scores are lower in the warning

condition compared to the control condition has been taken as evidence that the warning was effective in reducing faking.

In the second type of research design, laboratory research subjects are exposed to two manipulations. The first manipulation consists of the previously discussed warning versus control conditions. The second manipulation focuses on test taking instructions in which participants are asked to either “fake good” or to respond honestly in the personality test. It is believed that the “fake good” condition replicates the motivational state that actual job applicants are in when participating in high-stake testing situations. With this research design, the investigator is looking for a significant interaction between warning and the test taking instruction on test scores (for e.g., Goffin & Woods, 1995) In particular, the expectation is that there would be a relatively smaller test score difference between the “fake-good” and “honest take” conditions when a warning is present compared to when a warning is absent, i.e, reduced test score inflation. Evidence that supports this interaction has also been taken as evidence that the warning was effective in reducing faking.

Unfortunately, even though the two aforementioned research designs have been widely used, Dwight and Donovan (2003) have written that these research designs assume “*that the difference in mean scores is due to a reduction in faking* (pg. 2)”. The dearth of research on the assumed underlying process of reduced faking was what prompted this thesis. In other words, researchers should (but have yet to) consider whether a warning actually lowers faking, rather than test scores per se. Showing that a warning actually lowers faking is the first goal.

The second goal is to ensure that a warning does not do anything other than reduce faking. Specifically, Dwight and Donovan (2003) raised the possibility that some (honest) respondents could be motivated by a warning to engage in something other than reduced faking. That is, these respondents might potentially engage in "overcompensation", or behaviors that they believe will help them avoid being "identified" as having engaged in faking behavior. More specifically, to paraphrase Ellingson, Heggestad and Makarius (2012), an intervention "… should only work as a strategy for dealing with intentional distortion when individuals choose to respond more accurately… ". Thus, if a warning leads to anything else other than diminished faking then warning as a strategy to increase test score accuracy could actually backfire.

It should be clear by now that with the research designs that have been employed in the past, we cannot determine whether reduced test scores as a result of warning was entirely due to reduced faking only. In other words, the field needs to consider research designs such as mediation that explicitly models the underlying process. Simply put, until we actually examine whether diminished faking actually fully mediates the effect of warning on test score, we cannot be certain whether the effects of warning on test scores were transmitted via the sole mechanism of reduced faking only.

This proposal to examine mediation rather than main effects is also consistent with research on mediation (for e.g., see Hayes, 2009; Rucker, Preacher, Tormala, & Petty, 2011). Specifically, studies have shown that the lack of a statistically significant main effect does not necessarily provide any information regarding the presence of one or more statistically significant mediation effects. That is to say, the lack of support for a main effect (of warning on test scores) does not preclude the presence of mediation

effects. Restated, we should not summarily dismiss a warning as ineffective if it did not lower test scores (a main effect).

While I have presented an argument as to why we should discard the previous research designs and instead focus on two goals, showing that: (1) warning lowers faking, and (2) the effect of warning on test score is mediated by faking only, this is only part of the proposal. Specifically, in a bid to overcome the limitations that were present given the narrow definition of faking as test score inflation, I adopt a broader, more inclusive definition of faking. I then complement the two goals described above with two additional methods, to arrive at a multi-faceted test score validity approach to evaluate the effectiveness of warning as an intervention against faking.

A Multi-faceted Test Score Validity Approach

One key problem with the whole area research on faking is that there is no consensus regarding the definition of "faking". For example, in addition to test score inflation, as well as patterns of item response, other measures such as item response latencies have also been used to index the magnitude of faking (Holden & Hibbs, 1995; Vasilopoulos, Reilly, & Leaman, 2000). More recently, even eye-tracking (van Hooft & Born, 2012) has also been proposed as a way to detect faking.

As such, given the lack of an agreed upon definition of faking, it is perhaps understandable that the most easily obtained indicator among them all, a mean difference in observed personality test scores between subjects has been treated as *the* way to measure faking, rather than as one of the many possible indicators of faking (Burns & Christiansen, 2011). It is important to note that while various indices of faking have been

proposed, none have gone so far as to claim that any are each necessary and/or sufficient indicators of faking; merely possible indicators. To be more precise, while each measure may be consistent with the claims of faking being present, they do not rule out the role of other measures as simultaneously viable indicators of faking. This raises the question of what then constitutes an adequately comprehensive measure of faking.

Rather than debating which way is (ways are) best to measure faking, an alternative would be to treat faking as source(s) of systematic, irrelevant variance in test scores, i.e., as a threat to test score validity (Sackett, 2011). Adopting this definition would enable us to evaluate the effectiveness of a warning as a specific intervention against faking in a more comprehensive fashion, beyond just the magnitude of test score deflation that the warning causes. That is, if we consider faking as a threat to test score validity, then the effectiveness of a warning as an intervention against faking can be measured in terms of change in test score validity.

In this thesis, I defined validity as "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of a test (American Education Research Association [AERA], the American Psychological Association [APA], and the National Council on Measurement in Education [NCME], 1999, pg. 9)". This definition of validity implies that a test score itself is neither valid nor invalid. Instead, the issue of validity focuses on whether the *interpretation* of the test score can be justified or not. Applied to this thesis, claiming that a warning is an effective (*valid*) intervention against faking can only be supported to the extent that the evidence actually indicates that a warning improved test score validity via diminished faking only.

According to the Standards for Educational and Psychological Testing (Standards) published by AERA, APA and NCME, there are five different types of evidence that can be considered together to evaluate the validity of a given test score interpretation. That is, evidence for the overall validity of a test score interpretation can be drawn from the dimensions of: (1) test content, (2) the internal structure of the test, (3) the psychological processes used in test response, (4) the association among test scores and other variables, and (5) the consequences of test use. In this thesis, I focused on the role of warning on: (1) the internal structure of the test, (2) the psychological processes used in test response, and (3) the associations (correlations) among the personality test scores and other variables of interest.

Internal Structure of the Test

In terms of the "prevailing standard", past research designs have implicitly assumed the presence of measurement equivalence for personality scores across warning and no-warning conditions. Otherwise, the comparison of mean personality test scores across warned and no-warning conditions would not have made sense. Indeed, the assumption of measurement equivalence, i.e., an equivalent internal structure of the test across groups (e.g., warned and no-warning conditions), is a prerequisite for meaningful score comparisons across conditions (Vandenberg & Lance, 2000).

Recent research however has suggested that the assumption of measurement equivalence of personality test scores across warning and no-warning conditions may not be warranted. For example, DuVernet and colleagues (2014) found that a warning led to differential item functioning for some of the items in a personality test. This result suggests that at least some of the items in the personality test were not equivalent across

the warning and no-warning conditions. In other words, this highlights a need to actually verify whether the psychometric properties of the test score remain equivalent across the warning and no-warning conditions.

Assuming the presence of measurement equivalence when it is in fact absent has serious drawbacks. For example, if faking causes a loss of measurement equivalence, and if a warning is supposed to reduce faking, then any evidence indicating the absence of measurement equivalence could mean one of two things. First, it could mean that the warning failed to reduce faking and the associated loss of measurement equivalence that was caused by faking. Second, it could also mean that the warning itself might have caused the lack of measurement equivalence. Essentially, what it all boils down to is that it would only make sense to evaluate whether the warning was effective in raising test score validity in the presence of measurement equivalence. Restated, in the event that measurement equivalence is absent across warned and no-warning conditions, one could reasonably declare the warning to be an ineffective intervention against faking.

According to Meredith (1993) and others (e.g., Brown, 2006; Millsap, 2012), there are various levels of measurement equivalence. For the purposes of this thesis, we will need to consider only three levels. The three levels are configural, metric and scalar equivalence. First, when configural equivalence is present, this means that the pattern of personality test items that load onto each personality trait factor are the same, and that they lead to an equal number of factors overall across both warning and no-warning conditions (same form). Second, metric equivalence presumes configural equivalence and also requires the loadings for each item-factor relationship to be equivalent across conditions (same loading). Third, testing for scalar equivalence requires evidence for

metric equivalence before determining whether the intercepts of the item-factor loadings are the same across conditions (same intercept). The three levels of measurement equivalence can be tested using a multiple-group confirmatory factor analysis approach (for e.g., see Meade & Lautenschlager, 2004; Raju, Lafitte, & Byrne, 2002).

More importantly, demonstrating evidence for each level of measurement equivalence allows us to make certain interpretations that we would otherwise not be able to in the absence of measurement equivalence. That is to say, for the purposes of this thesis, demonstrating certain levels of measurement equivalence is a necessary pre-requisite before I can test other hypotheses regarding the impact of warning on test score validity. For example, evidence for configural equivalence is required before a personality trait score can even be calculated. In this case, using the same personality test items to arrive at a personality test score across the warning and no-warning conditions is only justified with configural equivalence. The absence of configural equivalence could mean that a personality test score is the result of up different items across conditions. If so, then the scores across conditions clearly refer to different constructs. In other words, configural equivalence is a necessary but insufficient condition for a direct comparison of test scores across conditions.

After demonstrating the presence of configural equivalence, testing for metric equivalence would ensure that the intervals between the units of measurement for each personality test item are equivalently scaled across conditions. Metric equivalence would then allow one to compare correlations involving the personality test scores across conditions. Assuming the presence of metric equivalence, evidence for scalar equivalence ensures that the test scores across conditions share the same origin (intercept). As such,

only with scalar equivalence can we then directly compare test scores across conditions. For example, to the extent that faking does lead to a lack of scalar equivalence, it makes "faked" and "honest" scores incomparable (for e.g., McFarland, Ryan, & Ellis, 2002; Schmit & Ryan, 1993).

It should be evident by now that assuming the presence of measurement equivalence when it is in fact absent exposes us to the risk of making unjustifiable or worse, incorrect claims. Indeed, one reason why faking continues to be considered a problem is because of the fear that it could affect the internal structure of the personality test score. For example, if a warning leads to a loss of configural equivalence, then it is clear that the personality construct have different meanings across the different warning conditions. In other words, a lack of configural equivalence would imply that a warning has a negative impact on the internal structure of the personality test. If so, then it would be straightforward to declare that a warning is harmful rather than helpful to test score validity.

Hypothesis 1: Warning leads to a loss of configural equivalence.

However, demonstrating configural equivalence is just the first step. Often we are not interested in just whether the personality test score has the same meaning across conditions. Usually, we are more interested in the similarity or differences in relationships among personality scores and other variables across conditions. To compare such correlations across conditions, evidence for metric equivalence is required. For example, without metric equivalence, we would not be able to evaluate whether a warning improved the predictive validity (strength of correlation) between personality test score and performance. Similarly, the absence of metric equivalence would preclude

a test of whether the effect of warning on test scores was mediated by reduced faking. This is because a mediation model is ultimately based on correlational data.

Hypothesis 2: Warning leads to a loss of metric equivalence.

Last but not least, metric equivalence is necessary but still insufficient to allow a direct comparison of test score means across conditions. To allow a direct a comparison of test score means across conditions, evidence for scalar equivalence is required. To be clear, the need for scalar equivalence is not just restricted to personality tests scores. For example, in addition to personality test scores, test users are often also interested in measures of applicant/test-taker reactions across warning and no-warning conditions (Burns et al., 2015; Converse et al., 2008; McFarland, 2003). That is to say, for such direct comparisons of any measure, such as personality and/or reaction measures to interventions across the warning and no-warning conditions to be valid, there has to be evidence for scalar equivalence for these measures.

Hypothesis 3: Warning leads to a loss of scalar equivalence.

Response Processes

First, it is obvious that an effective warning is one that reduces faking. While previous studies on warnings have used test score inflation as an index of faking, none have actually directly asked whether respondents faked, and how much, for which traits. The implicit assumption seems to be that no one will actually admit to faking. In contrast to previous studies on warning, I adopted the approach that some studies on faking have employed and directly asked whether respondents faked on the personality test (for e.g., Jansen, König, Stadelmann, & Kleinmann, 2012). In other words, if the warning has no

effect on faking, then I should not observe any differences in the participants' admissions to faking. However, if the warning does have an effect on faking, then there should be a difference.

Hypothesis 4: Faking will be lower in the warning compared to the no warning condition.

In this thesis, I am not just interested in whether a warning has the desired effect on faking. I am also interested in whether the psychological processes that respondents *actually* engage in, is consistent with psychological processes we *expect* them to engage in; when they receive a warning (Bornstein, 2011; Borsboom, Mellenbergh, & van Heerden, 2004). For example, warnings have been presented as messages to the test taker before they begin the personality test. While the content of the warning message may vary (Burns et al., 2015; Pace & Borman, 2006), research suggests that warnings appear to be most effective in lowering test scores when the warning includes information stating that faking behavior can be identified and that the test taker will have to bear the negative consequences once faking has been detected (Dwight & Donovan, 2003). As such, if the aim of the warning is to lower faking behavior, then it should be a reasonable to hypothesize that a warning should first lower the motivation to fake (Ellingson & McFarland, 2011, Marcus, 2009; McFarland & Ryan, 2006).

Hypothesis 5: Motivation to fake will be lower in the warning compared to the no warning condition.

Nevertheless, as mentioned previously, while a warning may lead to lower motivation to fake, this need not necessarily translate to actually lower faking behavior.

For example, a lower motivation to fake could perhaps also lead to more "overcompensation" instead. More specifically, a warning could "cause a respondent to overcompensate in their responding so as to ensure that they are not identified as faking" (Dwight & Donovan, 2003, pg. 5). In other words, showing that a warning leads to lower motivation to fake would be necessary but insufficient to support the claim that a warning was effective in reducing faking. Put differently, in addition to showing that a warning lowers motivation to fake, it would also be necessary to establish a link between motivation to fake and faking behavior as well.

Hypothesis 6: Motivation to fake is positively associated with faking.

More importantly, to the extent that a warning does have an effect on test scores, the effect of a warning on test scores should also be serially mediated, first via lower motivation to fake, and then followed by lower faking (see model in Figure 1).

Hypothesis 7: The effect of warning on test scores is mediated by motivation to fake and faking.

Personality Test Criterion-Related Validity

Last but not least, in addition to examining the effects of a warning on test score measurement equivalence and response processes, test users are also often interested in whether warning improves the predictive validity of the personality test scores. For example, simulations of faking behavior have shown that faking behavior can harm the predictive validity of personality test scores, which affects who gets hired (Berry & Sackett, 2009; Converse, Peterson, & Griffith, 2009; Donovan, Dwight, & Schnieder, 2014; Komar, Brown, Komar, & Robie, 2008). Even in a more complex situation, where

multiple predictors other than just personality tests were used, faking in the personality test harms predictive validity and detracts from justifiable selection decisions (Converse et al., 2009). To put it simply, another way to evaluate whether a warning is an effective intervention against faking is by examining whether it improves the capability of personality test scores to predict other variables of interest.

Hypothesis 8: A warning moderates the magnitude of a personality trait-criterion relationship, such that the strength of the relationship is higher in the presence of a warning.

In summary, I have argued that the designs employed in previous research are insufficient for determining whether warning is an effective means for producing more valid personality test scores. As an alternative to the previous research designs, I adopted a test score validity approach and examined the effects of warning along three dimensions of test score validity. First, does a warning affect the internal structure of the test score? Assuming a warning does not lead to a lack of measurement, does warning only work through the expected mediation process? Finally, assuming the evidence supports the previous two hypotheses, does warning actually improve the predictive validity of personality test scores?

Examining the effects of warning along these three dimensions of test score validity are consistent with the *1999 Standards for Educational and Psychological Testing* (Standards, 1999), which calls for the examination of both intended and potentially unintended consequences of test (in this case, warning) use. In our quest to reduce faking, we should strive to ensure that any deliberate intervention that we apply, such as a warning, must first do no harm (i.e., *primum non nocere*) to the validity of the

test scores. In the first study, I will attempt to replicate previous results (e.g., Adair, 2014) by examining whether warning lowers personality test scores (aka “prevailing standard”) as well as test whether warning actually lowers faking behavior (Hypothesis 4).

Chapter 2: Study 1

Method

Participants and Research Design

Study 1 was a two (Incentive) by two (Warning) completely randomized factorial ANOVA. A total of 250 participants were recruited online from Amazon Mechanical Turk. However, after reviewing participant responses to determine whether they were paying sufficient attention to the survey, a total of 60 participants were dropped. Thus, the effective sample size was 190.

In terms of gender, 102 (53.7%) were female, 86 (45.2%) were male, and 2 (1%) did not report their gender. Participants' age ranged from 19 to 74 (M = 39.30, SD = 12.79). More than two thirds of the sample were either self-employed (13.2%) or employed in an organization (57.9%). The remainder was either unemployed (16.3%), students (3.7%), retirees (5.3%) or could not be categorized into any of the categories above (3.7%). Their working experience ranged from 0 to 55 years (M = 17.80, SD = 11.83).

Manipulations

*Incentive***.** Participants were randomly assigned to either an incentive or no incentive condition. In the incentive condition, they were told that the participant with the highest personality test score would receive a $20 bonus. Participants in the no incentive condition did not receive this message.

Warning. Participants in the warning condition were provided with a message (see Appendix 1) before they took the personality test. Participants in the no warning condition did not receive this message.

Procedure

Participants were told that they would be assuming the role of an applicant for a retail salesperson position. To ensure that all participants understood what the job position entailed, they were provided with a relevant job description that was obtained from O*Net Online (see Appendix 2). This job description ranked the importance of various work styles (i.e., manifestations of personality traits) that are desirable for the targeted job.

Participants were then randomly assigned to one of the four experimental conditions. That is, participants either received: (1) the incentive and warning messages, (2) the incentive message only, (3) the warning message only, or (4) neither the incentive nor the warning message. After receiving the experimental manipulation, they then took the personality test. After completing the personality test, participants were asked to complete a set of statements regarding their behavior in the personality test. These statements were specifically intended to measure participants' level of self-reported faking behavior on the personality test. After this, participants were fully debriefed.

Measures

Personality. I used the HEXACO-60 (Ashton & Lee, 2009) as the personality test. Responses to the HEXACO-60 (see Appendix 3) were measured using a 5-point agreement scale (1 = Strongly Disagree to 5 = Strongly Agree). Each of the six

HEXACO facet scores was calculated by computing the mean of the ten items that belonged to each facet. In my sample, the Cronbach's alphas for the six HEXACO facet scores were all above 0.75 (see Table 1).

Faking. I measured participants' faking behavior in the personality test with a series of fourteen statements (see Appendix 4). Specifically, participants responded either "True" (coded 1) or "False" (coded 0) to statements such as: "I exaggerated my responses to make myself look better than I actually am" and "I tended to de-emphasize or "play down" what some might consider my negative attributes". I tried to encourage honest responses to these faking behavior statements by using the Randomized Response Technique (RRT). The RRT has been applied in previous research to gather sensitive information from participants (for e.g., see Donovan, Dwight, & Hurtz, 2003; König, Hafsteinsson, Jansen, & Stadelmann, 2011: König, Wong, & Cen, 2012). The RRT works by asking participants to follow one of two rules when responding to the statements. The first rule asks participants to always honestly answer the statement whereas the second rule asks them to always respond "True" regardless of whether that was in fact the truthful response or not. The rule to be used for each statement is randomly determined.

In the present study I enacted the RRT by generating a random number sequence for each participant. Participants chose a number from that random number sequence before answering a statement. If their chosen number was even, they were told to respond honestly to the statement. If their chosen number was odd, they were told to respond "True" regardless of what the statement said. After responding to a statement, participants crossed off that number from the sequence and chose another number from the remaining numbers in the random number sequence for the next statement. This

process was repeated until all the statements were answered. The faking statements were scored so that a “True” response received a score of 1.

It has been shown that the RRT guarantees the anonymity of individual responses. In other words, while the RRT approach encourages honest responses to sensitive items, the trade-off is that it is not possible to compute individual level scores for faking behavior. As a result comparisons of faking behavior obtained using the RRT can only be interpreted at the aggregate (i.e., between experimental condition) level of analysis. The estimated proportions of honest responses for each faking behavior statement in each experimental condition were averaged within each condition. This gave rise to four faking behavior frequency estimates, i.e., one per condition (see Table 2).

Results

Means, standard deviations, and inter-correlations for the Study 1 variables are presented in Table 1. In a bid to retain comparability with past studies on warning, I replicated the research design employed in past studies and subjected the six HEXACO trait scores to a completely randomized 2 by 2 factorial ANOVA. First, I did not find any statistically significant main effects for the incentive manipulation on the HEXACO trait scores (see Table 3). However, I did find a statistically significant main effect for the warning manipulation for one of the six HEXACO trait scores (Table 4). Specifically, I found that warning had a significant effect on Agreeableness, $F(1, 188) = 9.21$, $p < .05$, $d = -.44$, such that the mean Agreeableness score was lower in the warning ($M = 3.29$, $SD = .76$) compared to the no warning condition ($M = 3.63$, $SD = .78$). In addition, the warning manipulation also had a main effect that was trending towards significance for Extraversion, $F(1, 188) = 3.52$, $p < .10$, $d = -.32$. In this case, Extraversion scores

appeared to be lower in the warning (M = 3.35, SD = .76) compared to the no warning condition (M = 3.57, SD = .76).

With regard to the interaction between incentive and warning on HEXACO test scores, I did not find any statistically significant interaction effects (see Tables 5). The only interaction effect that was trending towards significance was for Openness to Experience, $F(1, 186) = 2.79$, $p < .10$, partial $\eta^2 = .015$. Examination of the simple main effects suggested that warning tended to lower Openness scores in the presence of an incentive (d = -.37), but not in the absence of an incentive (d = .10). Overall, the results replicated past research suggesting that the warning manipulation used in this study successfully reduced personality test scores.

I then examined whether warning lowered self-reported faking using pair-wise z-tests. First, across all four experimental conditions, the level of self-reported faking was all statistically significantly different from zero (see Table 2). This indicated that participants admitted to some level of faking in the personality test in all of the four experimental conditions. In particular, faking was also statistically significantly different from zero even in the no incentive, no warning condition.

Next, I examined whether the warning manipulation decreased self-reported faking behavior on the personality test, again using pairwise z-tests. Surprisingly, the difference in the level of faking behavior was not statistically significant significantly between the no incentive and no warning condition (the control, i.e., no intervention condition) and any of the other three conditions (all ps > .10). In other words, Hypothesis 4 was not supported. There was no evidence that the warning decreased faking.

Study 1 Discussion

Overall, Study 1 demonstrated that I was able to replicate the typically obtained effect of the warning manipulation using a Mechanical Turk sample. Specifically, the warning lowered Agreeableness scores. However, while Study 1 replicated previous findings in terms of warning and lower personality test scores, there was no support for Hypothesis 4. That is, I could not reject the null hypothesis that the warning led to less faking. This result was surprising because the main effect of warning on Agreeableness score as well as trending effects for Extraversion and Openness were in the expected direction. In other words, it is troubling that lower test scores in the warning condition were not observed together with lower levels of faking even with the the RRT design. Remember that the RRT design was supposed to increases participants' belief that their responses to sensitive questions, such as admitting to faking, are anonymous. As such, failure to reject the null for Hypothesis 4 raises questions regarding the extent that the observed reduction in personality test scores can be attributed to a reduction in faking behavior.

In other words, the apparent lack of support for Hypothesis 4 indicates a potential problem for the use of warning as an intervention against faking. This is because the primary justification for using warnings in a high stakes testing context is that we expect the warning to lead to reduced faking, i.e., more honest responses from job applicants. Stakeholders certainly did not expect a warning to lower test scores without reducing faking as well. As such, Study 1 raises questions about the defensibility of using warnings to increase test score validity in high-stake testing situations. Therefore, this result provides initial support for my claim that previous research designs, i.e., the

“prevailing standard” which simply focuses on whether a warning reduces personality test scores, are insufficient to evaluate whether a warning is a valid intervention against faking.

Nevertheless, there could be alternative explanations that could potentially explain the lack of support for Hypothesis 4. For example, how I measured faking in Study1 could have mattered. Specifically, the faking behavior items were written to measure faking behavior at a gross level (see Appendix 4). That is, I assessed the magnitude of faking over the entire personality test instead of the degree to which people fake on particular personality traits or personality test items. In other words, the faking items may not have captured any reduced faking that was limited to certain personality traits.

It is also possible that providing participants with the job description (see Appendix 2) could have activated individual schemas for the perfect retail sales employee. That is, participants could have differentially faked the personality test as a function of how they perceived the centrality of a given personality dimension (McFarland & Ryan, 2000). Therefore, it is possible that the effect of warning on the personality traits with lower scores in the warning condition may not have been adequately captured with the faking behavior items used.

At the same time, the design employed in Study 1 cannot rule out the claim that the warning might have caused systematic but irrelevant effects - such as “overcompensation” by otherwise honest participants who do not wish to appear to be identified as fakers (e.g., Dwight & Donovan. 2003). As such, I conducted a second study

to address these limitations. In the second study, I used different items to measure faking behavior and dropped the RRT methodology.

First, I re-wrote the faking behavior items so that they focused on particular personality traits. Second, I dropped the RRT because of the limitations it imposes on hypothesis testing. While the RRT has its advantages, it also requires larger samples and suffers from data attrition when participants fail to comply with the RRT instructions. More importantly, because the RRT limits the comparisons of faking to between experimental conditions only, I could not test whether faking was a mediator of the warning-test score relationship. Admittedly, absent the RRT methodology, participants might under-report their level of faking behavior. However, given that Study 1 found that people were admitting to non-negligible levels of faking, even in the control (i.e., no incentive and no-warning) condition, it seems appropriate to focus on whether the warning indeed acts via hypothesized mediating role of (diminished) faking.

In summary, in Study 2 I examined whether a warning reduced: test scores ("the prevailing standard"), faking (Hypothesis 4) and motivation to fake (Hypothesis 5). I also examined whether motivation to fake was positively associated with faking (Hypothesis 6) as well as the hypothesized mediating roles of motivation to fake and faking (Hypothesis 7). Lastly, I also included a pre-warning manipulation measure of personality. This allowed me to evaluate whether the warning improved the strength of the correlation between the pre and post manipulation measures of personality (Hypothesis 8).

Chapter 3: Study 2

Method

Participants and Research Design

Study 2 was a completely randomized block design with two between person factors (i.e., incentive & warning) and one within person factor (repeated measurement of personality). A total of 140 participants were initially recruited from Amazon Mechanical Turk, however, 26 participants were dropped after performing an attention check on the data. The effective sample size was 116. Of these participants, 66 were female (56.9%), 40 were male (34.5%) and 10 did not report gender (8.6%). The age of participants who were retained ranged from 20 to 68 (M = 37.07, SD = 11.81). More than two thirds of the sample (71.7%) were either self-employed (N = 16) or employed in an organization (N = 60). The rest were unemployed (N = 14), students (N = 8), retired (N = 4) or could not be categorized into any of the former categories (N = 4). For years of working experience, participants reported values ranging from 0 to 50 years (M = 16.05, SD = 10.81).

Manipulations

*Incentive***.** Participants were randomly assigned to either an incentive or no incentive condition. In the incentive condition, they were told that that the top five personality scorers would receive $10. Participants in the no incentive condition did not receive this message.

Warning. The same manipulation used in Study 1 was used in this study.

Procedure

The procedure was the same as Study 1 except that before receiving the retail sales job description, participants were provided with the mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006), ostensibly as “practice” personality test items. After completing the mini-IPIP, they were randomly assigned to one of the 4 experimental conditions. Participants then completed a subset of items from the HEXACO-60 personality test that corresponded to four of the traits measured by the mini-IPIP items, followed by items measuring motivation to fake and faking behavior on the HEXACO personality test. All participants were thoroughly debriefed.

Measures

Pre-manipulation measure of Personality. Participants completed twenty items from the mini-IPIP (Donnellan et al., 2006) using a 5-point Likert scale (1 = Very Inaccurate to 5 = Very Accurate), specifically to measure the four traits: (1) Extraversion, (2) Agreeableness, (3) Conscientiousness and (4) Neuroticism and (5) Intellect/Imagination (see Appendix 5). Each mini-IPIP trait was calculated by computing the mean of the items belonging to each trait. The internal consistency reliability for the mini-IPIP traits was at least 0.77 or higher (see Table 7).

Post manipulation measure of Personality. Participants completed a subset of forty items from the HEXACO-60 (Ashton & Lee, 2009) using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Specifically, four traits were measured, including: (1) Extraversion, (2) Agreeableness, (3) Conscientiousness and (4) Emotionality. Each HEXACO trait score was calculated by computing the mean of the items belonging to each facet. The internal consistency reliability for the HEXACO facets was at least 0.80 or higher (see Table 7).

Motivation to Fake. Participants responded to three items using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) regarding their motivation to fake on the HEXACO personality test. The items were: “I was motivated to get a top score”, “Getting a top score was important to me”, and “I did not really care about how well I presented myself” (reverse scored). A composite motivation to fake score was calculated by computing the mean of the 3 items (α = .73). Higher values indicate greater motivation to fake.

Faking. Participants completed eight items (see Appendix 6) using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) regarding their responses on the HEXACO personality test. The level of faking behavior was calculated by computing the mean across the eight items (α = .77). Higher values indicate greater faking.

Results

Means, standard deviations, and inter-correlations are presented in Table 7. I first tested whether the incentive and warning manipulations had an effect on the mini-IPIP. In line with my expectations and the fact that the mini-IPIP was administered before the experimental manipulations, I did not find any main effect of incentive (see Table 8) or warning (see Table 9) on any of the mini-IPIP measures. I also did not find any statistically significant interaction effects (see Tables 10 and 11). This finding supports that the random assignment of participants to conditions worked.

Next, I performed an ANOVA for the four HEXACO measures. This analysis showed that there was neither a main effect of incentive (Table 12) nor a main effect of warning (Table 13) on any of the four HEXACO traits. I also did not find any statistically

significant interaction effects between incentive and warning on any of the four HEXACO traits (Tables 14 and 15). If I adopted the "prevailing standard", I would have stopped here and declare that the warning failed to reduce faking.

However, because this thesis is interested in whether the warning increases test score validity rather than simply lowering test scores, I also examined the effects of the incentive and warning manipulations on motivation to fake and faking. First, an ANOVA showed that incentive did not have an effect on motivation to fake, $F(1, 106) = .00$, $p >. 10$, $d = .00$ or faking, $F(1, 105) = .75$, $p >. 10$, $d = -.17$ (Table 16). Second, while warning did not have a main effect on motivation to fake, $F(1, 106) = .03$, $p >. 10$, $d = -.02$, it did have a main effect on faking behavior, $F(1, 105) = 5.78$, $p < . 05$, $d = -.45$ (Table 17). Specifically, there was less faking in the warning ($M = 1.74$, $SD = .66$) compared to the no-warning ($M = 2.08$, $SD = .82$) condition. In addition, there were no interaction effects between incentive and warning for motivation to fake, $F(1, 104) = .08$, $p > . 10$, partial $\eta^2 = .00$ (Table 18), or for faking, $F(1, 103) = .18$, $p > . 10$, partial $\eta^2 = .00$ (Table 18 and 19). In other words, the claim that warning lowers motivation to fake (Hypothesis 5) was not supported, but the claim that warning lowers faking (Hypothesis 4) was.

Hypothesis 6 predicted that motivation to fake would be positively associated with faking. Contrary to this hypothesis, there was no evidence to support the claim that motivation to fake was positively associated with faking, $r(104) = -.04$, $p > .10$ (Table 7). I then tested Hypothesis 7, which predicted that the effects of warning on personality test scores would be mediated by motivation to fake and faking (see Figure 1). To test Hypothesis 7, I estimated a mediation model using the PROCESS macro (model type = 6) in SPSS (Hayes, 2013) for each of the four HEXACO traits. Estimates for the

mediation models were based on 10,000 bootstrap samples and were summarized in Table 20.

Inspection of Table 20 shows that for each mediation model, there are a total of six effects. The first four columns of this table provide the estimates of three specific indirect effects. The first specific indirect effect represents a typical simple mediation model, which in this case represents the effect of warning on test score via motivation to fake ($X \rightarrow M_1 \rightarrow Y$). This specific indirect effect was estimated with a_1b_1. The second specific indirect effect represents a two-stage serial mediation effect, where the effect of warning on test score is first transmitted via motivation to fake, followed by faking ($X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$). This effect was estimated with $a_1d_{21}b_2$. The third specific indirect effect represents the second simple mediation effect, where the effect of warning on test score occurs via faking ($X \rightarrow M_2 \rightarrow Y$). This was estimated with a_2b_2.

The fourth column in Table 20, is simply the sum of the previous three specific indirect effects, i.e., the total indirect effect. The sum of the indirect was included to highlight the importance of specifying and testing for specific indirect effects. This is because the total indirect effect depends not just on the size but also the sign of its composite specific indirect effects. In other words, in cases when there is a “zero” total indirect effect, failure to inspect the specific indirect effects can lead to an incorrect conclusion that all of the specific indirect effects are also zero. While this instance is certainly possible, other instances are possible as well. For example, two specific indirect effects of opposite but equal magnitude could sum also to an overall “zero” total indirect effect. Apart from the total indirect effect, the remaining portion of the total effect is attributed to the direct effect, and this was estimated by c^1, represented in the fifth

column in Table 20. The sixth and last column simply represents the total cumulative effect of the warning on the test score.

The regression analyses for each of the mediation models for Extraversion, Agreeableness, Conscientiousness and Emotionality can be found in Tables 21, 22, 23 and 24 respectively. While I also conducted a separate set of analyses, which included the pre-manipulation measure of personality as a covariate for the dependent variable, the post-manipulation measure of personality, the results were effectively identical to the model without the covariate. Thus, I only reported the set of results without the covariate for parsimony.

Inspection of the specific indirect effect estimates in Table 20 showed that none of them were statistically significant. This was because none of the 95% confidence intervals for the specific indirect effect estimates excluded zero. In addition, none of the total indirect effects were statistically significant as well. In other words, there was no evidence that motivation to fake or faking mediated the effects of warning on test score, i.e., Hypothesis 7 was not supported.

I then examined whether the warning increased the strength of association between the pre- and post-manipulation measures of personality (Hypothesis 8). A summary of these correlations in each of the four experimental conditions can be found in Table 25. The separate multiple moderated regression (MMR) models that tested for the presence of interaction effects were reported in Tables 26, 27, 28, and 29 respectively. Inspection of the Table 26 showed that there was no evidence of a 3-way interaction among pre-manipulation Extraversion, incentive and warning on post-manipulation Extraversion, $b_7 = .01$, 95% CI [-.14, .17], $se_{b7} = .08$, $t(101) = .19$, $p > .10$. Specifically,

including the 3-way interaction term in Model 3 did not lead to a statistically significant ΔR^2 = .00, compared to Model 2, $\Delta F(1, 101) = .04$, $p > .10$. There was also no evidence for any 2-way interactions. Specifically, including the 2-way interaction terms in Model 2 did not lead to a significant ΔR^2 = = .01, compared to Model 1, $\Delta F(3, 102) = .21$, $p > .10$. In other words, I failed to reject the null hypothesis of equivalent correlations between the pre and post manipulation measures of Extraversion across the no-warning and warning conditions.

Inspection of Table 27 also showed no evidence of a 3-way interaction for Agreeableness, b_7 = -.08 [-.25, .09], se_{b7} = .09, $t(102) = -.94$, $p > .10$. Specifically, the ΔR^2 = .01 after including the 3-way interaction term was not statistically significant, $\Delta F(1, 102) = .88$, $p > .10$. There was also no evidence for any 2-way interactions, $\Delta F(3, 103) = .22$, $p > .10$, ΔR^2 = .01. In other words, there was no evidence that the warning improved the pre- and post- correlation for Agreeableness.

Inspection of Table 28 similarly showed no evidence of a 3-way interaction for Conscientiousness, b_7 = .08 [-.04, .20], se_{b7} = .06, $t(101) = 1.30$, $p > .10$; ΔR^2 = .01, $\Delta F(1, 101) = 1.69$, $p > .10$. There was also no evidence for any 2-way interactions, $\Delta F(3, 103) = .38$, $p > .10$, ΔR^2 = .01. Restated, there was no indication that the warning led to a stronger association between pre- and post-manipulation measures of Conscientiousness.

Finally, inspection of Table 29 did not indicate the presence for a 3-way interaction for Emotionality, b_7 = -.09 [-.22, .04], se_{b7} = .07, $t(101) = -1.40$, $p > .10$; $\Delta F(1, 101) = 1.97$, $p > .10$, ΔR^2 = .02. The results also did not indicate the presence for any 2-way interactions, $\Delta F(3, 102) = 1.09$, $p > .10$, ΔR^2 = .03. Thus Hypothesis 8 was not supported for any of the four HEXACO traits.

Study 2 Discussion

Overall, in Study 2, there was support for Hypothesis 4 (warning lowers faking), but not for the rest of the Hypotheses (5, 6, 7 and 8) that were tested. Nevertheless, there are two salient findings in Study 2. The first salient point is that despite a failure to reject the null for equal test scores across conditions, the null was rejected for equal faking across conditions in the expected direction. Restated, Study 2 showed a pattern of results which highlights the inadequacy of the "prevailing standard" to evaluate the effectiveness of a warning. Put differently, had we neglected to test Hypothesis 4 and instead only limited ourselves to examining whether the warning lowered test scores or not (the "prevailing standard"), we would have arrived at a different conclusion, which is that the warning failed to reduce faking in Study 2. The second salient point is that faking did not mediate the effects of warning on test scores. To the extent that the warning had the desired effect on faking, then the lack of support for a mediation result suggests that an effective warning could potentially reduce faking without affecting test scores.

At the same time, probing the underlying process model did reveal some unexpected findings as well. Specifically, there was also no evidence that the warning had an impact on motivation to fake, or that motivation to fake was positively associated with faking. More importantly, there was neither a simple nor serial mediation effect for the effect of warning on test score via motivation to fake. These findings were all the more surprising because motivation has been considered an integral part of theories on faking (Ellingson & McFarland, 2011; Snell, Sydell, & Lueke, 1999, Tett & Simonet, 2011). To be clear, it is hard to reconcile the finding that systematic differences in faking had occurred without a motivation component.

One way to reconcile this conflicting pattern of result is to focus on the apparent inconsistency. For example, some could argue that the “conflict” exists because of the sole statistically significant finding in Study 2. Restated, if the statistically significant main effect of warning on faking is actually a Type 1 error, i.e., an incorrect rejection of the null, then the overall pattern of results could make sense. For example, inspection of the faking measure across warning and no-warning conditions did reveal a statistically significant difference, indicating that respondents had disagreed more strongly with the faking statements in the warning compared to the no-warning condition. As such, some could argue that disagreeing more strongly should not be treated as an indication of substantive differences in faking, i.e, there was no change in “sign” (from agreeing to disagreeing). If we were to accept this line of argument, then the overall pattern would make sense. Put simply, then one could also argue that the warning was effectively inert (the main effect of warning on faking should be neglected), and did not have any effect at all in Study 2.

In other words, I have offered two ways to explain the effects of warning in Study 2. To reiterate, the first account tentatively indicates that the warning reduced faking without necessarily reducing either the motivation to fake or the test scores. Keeping in mind that faking can manifest itself in ways other than changes in the test score mean (Burns & Christansen, 2011), this first account would support my claims the “prevailing standard” needs to be reconsidered. In contrast, the second account remains ambiguous regarding the usefulness of the “prevailing standard” because of the lack of any substantively significant findings.

Nevertheless, Study 2 highlights the importance of verifying *how* the warning works. So far, Studies 1 and 2 demonstrated that I could fruitfully examine the hypothesized effects of the warning manipulation using samples from the Amazon Mechanical Turk population. Simply out, the data has suggested that it is indeed possible to do so. However, given at least two potential ways to interpret the results in Study 2, it still remains unclear whether the warning truly reduced faking and if so, if the effects of warning on test scores acted solely via the expected mechanisms only. As such, to investigate whether the warning indeed works in the way we expect it to, I conducted a third study to simulate a high stakes job applicant context as closely as possible.

Thus in Study 3, I told all participants that they could receive a “hiring bonus” ($10) contingent on their performance on the personality test. I also replaced the pre-manipulation measure of personality with alternative measures of impression management, organizational citizenship behavior and counter-productive work behavior. All eight Hypotheses were examined in Study 3.

Chapter 4: Study 3

Method

Participants and Research Design

Study 3 was a one-way completely randomized ANOVA. A total of 719 participants were recruited online from Amazon Mechanical Turk. However, after reviewing participant responses to determine whether they were paying sufficient attention to the survey, a total of 99 participants were dropped. Thus, the effective sample size was 620.

In terms of gender, 387 (62.4%) were female, 195 (31.5%) were male, and 38 (6.1%) did not report their gender. Participants' age ranged from 12 to 74 (M = 36.86, SD = 12.59). In terms of employment status, 107 (17.3%) were self-employed, 341 (55.0%) were employed in an organization. There were 64 (10.3%) unemployed, 43 (6.9%) students, 17 retirees (2.7%) and 13 (2.1%) who could not be categorized into any of the categories above. Their working experience ranged from 0 to 57 years (M = 15.65, SD = 11.85).

In terms of research design, prior to the experimental manipulation, participants first completed three scales that measured (1) organizational citizenship behavior, (2) counterproductive work behavior and (3) impression management. I then randomly assigned participants to either the warning or no warning (no intervention) condition. The warning intervention used in Study 3 was the same as the one used in Studies 1 and 2. All the participants were told that the top 10% (for e.g., 60 out of 600) of applicants who were going to be "hired", based on their personality test scores, would receive a $10

bonus. They were presented with the same job analysis report as per in Studies 1 and 2 and completed a personality test. After participants completed the personality test, they were asked to respond to a series of statements regarding their responses of the personality test. These items were then used to calculate the (1) motivation to fake and (2) faking measures.

Measures

Unless specified otherwise, participants responded to the items in the following measures using a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree).

Organizational Citizenship Behavior (OCB). (α = .89) Participants completed ten items (see Appendix 7) of the OCB-Checklist (Spector, Bauer, & Fox, 2010). The items were measured using a 5-point frequency scale (1 = Never to 5 = Everyday). The OCB score was calculated by summing the responses across the ten items. Higher values indicate a higher level of OCB.

*Counterproductive Work Behavior (CWB)***.** (α = .84) Participants completed ten items (see Appendix 8) of the CWB-Checklist (Spector et al., 2010) measuring counterproductive work behavior, with a 5-point frequency scale (1 = Never to 5 = Everyday). The CWB score was calculated by summing the responses across the ten items. Higher values indicate a higher level of CWB.

Impression Management (IM). (α = .84) Participants completed twenty items (see Appendix 9) of the Impression Management (IM) subscale of the Balanced Inventory of Desirable Responding (BIDR, Paulhus, 1984). Items were anchored with a 7-point accuracy scale (1 = Not True to 7 = Very True). Responses that were either a 6 or a 7 on

the 7-point accuracy scale were recoded as a 1 and all other responses (1 through 5), were recoded as a zero. The individual level IM score was then calculated by summing up the number of responses that were recoded as a 1 across the twenty items. Higher values indicate a higher level of IM.

Personality. The HEXACO-60 (Ashton & Lee, 2009) was used as the personality test (see Appendix 3). Each of the six HEXACO trait scores was calculated by computing the mean of the ten items belonging to each trait. The Cronbach's alpha for each of the six traits were at least .76 or higher (see Table 30).

Motivation to Fake. (α = .79) I used the same three items from Study 2 to measure motivation to fake.

Faking. (α = .87) I used the same eight items from Study 2 to measure faking behavior.

Results

Means, standard deviations, and inter-correlations are presented in Table 30. I first tested whether the warning manipulation had an effect on the pre-manipulation measures. In line with my expectations, I did not find any main effect of warning on any of the three pre-manipulation measures of impression management, organizational citizenship behavior or counterproductive work behavior (Table 31). This finding supports that the random assignment of participants to conditions worked.

Next, to ensure that the interpretations of the results for the subsequent hypotheses could be justified, I tested whether there was configural, metric and scalar equivalence across warning and no warning conditions. For example, to test for

mediation (Hypothesis 7) as well as to compare the correlations between the pre- and post-manipulation measures across the warning and no-warning conditions (Hypothesis 8), there needs to be at least metric equivalence. In addition, direct comparisons of personality test scores, motivation to fake (Hypothesis 5) and faking (Hypothesis 4) across warning and no-warning conditions requires evidence for scalar equivalence to be valid.

Evidence for measurement equivalence (Hypotheses 1, 2, and 3) was examined using multi-group confirmatory factor analysis (MG-CFA), model fit and model comparisons. All MG-CFA models were all estimated using the lavaan package (Rossel, 2012) in R (R Core Team, 2014, see Appendix 10 for the code that was used). To assess goodness of model fit, other than the traditional Chi-square statistic, I also relied on four other measures of model fit. They were: (1) the Comparative Fit Index (Bentler, 1990), (2) the Tucker-Lewis Index (TLI), which is also referred to as the Bentler-Bonett Non-Normed Fit Index (NNFI; Bentler & Bonett, 1980), (3) the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), and (4) the Standardized Root Mean Square Residual (SRMR; Bentler, 1995). I included these goodness-of-fit measures based on the extensive simulation evidence provided by Hu and Bentler (1999). Specifically, Hu and Bentler also proposed that values of .95 or greater for CFI and TLI indicated good fit. In contrast, for RMSEA and SRMR, values less than .06 and .08 respectively indicated good fit. All model fit indices are reported in Table 32.

For model comparisons, I considered two criteria. That is, in addition to using the traditional Chi-square difference statistic to decide which model demonstrated a comparatively better fit, I also considered the difference in Comparative Fit Index

(ΔCFI). First, this was because the traditional Chi-square difference statistic has been shown to be sensitive to sample size (Brannick, 1995; Kelloway, 1995). Second, a recent simulation study by Cheung and Rensvold (2002) found that ΔCFI was a reliable and robust measure of model fit. Specifically, if the difference (loss) in CFI between models is less than .01, then it would indicate that the fit for the more constrained model was acceptable. Model comparison results are reported in Table 33.

Before testing Hypotheses 1, 2, and 3, I first estimated an initial confirmatory factor analysis (CFA) model that specified eight factors in the data as the baseline (Model 0). The eight factors included the six HEXACO traits, motivation to fake and faking. The following fit indices were observed for the baseline (Model 0): $\chi^2(2386) = 37286.52$, $p < .05$, CFI = .79, TLI = .78, RMSEA = .15, SRMR = .02.

Next, to test Hypothesis 1, I estimated a multi-group confirmatory factor analysis (MG-CFA) model that specified equal form for the eight factors across the warning and no-warning conditions (Model 1, Table 32). Adding the first constraint of equal form led to the following fit indices: $\chi^2(4772) = 9275.49$, $p < .05$, CFI = .70, TLI = .69, RMSEA = .06, SRMR = .08. While the CFI and TLI values for Model 1 were less than ideal (< .95), the values for RMSEA and SRMR suggested that the model demonstrated good fit. Overall, this suggested that the warning did not lead to a loss of configural equivalence, i.e., Hypothesis 1 was not supported.

To test Hypothesis 2, I estimated second MG-CFA model (Model 2, Table 32) that was based on Model 1 but with the additional second constraint of equal item-factor loadings across warning and no-warning conditions. This led to the following fit indices for Model 2: $\chi^2(4835) = 9365.30$, $p < .05$, CFI = .70, TLI = .69, RMSEA = .06, SRMR =

.08. While the Chi-square difference test was significant, $\chi^2(63) = 89.81$, $p < .05$, the difference in CFI was less than .01 (see Table 33). In other words, based on the difference in CFI criterion (Cheung & Rensvold, 2002), this suggested that metric equivalence was present across warning and no-warning conditions, i.e., Hypothesis 2 was not supported.

To test Hypothesis 3, I estimated a third MG-CFA model (Model 3, Table 32) that was based on Model 2 but with a third constraint of equal item intercepts across the warning and no-warning conditions. This led to the following fit indices for Model 3: $\chi^2(4898) = 9427.77$, $p < .05$, CFI = .70, TLI = .70, RMSEA = .06, SRMR = .08. In addition, because the Chi-square difference test was not significant, $\chi^2(63) = 62.47$, $p > .05$, and the difference in CFI was less also than .01 (Table 33), scalar equivalence appears to be present across warning and no-warning conditions, i.e., Hypothesis 3 was not supported.

After probing for and establishing some evidence for the presence of configural, metric and scalar equivalence for the HEXACO, motivation to fake and faking measures, I proceeded to directly compare the personality test scores (aka “prevailing standard”), motivation to fake (Hypothesis 5), and faking (Hypothesis 4), across warning and no-warning conditions.

To examine whether the warning had an effect of personality test scores, I ran an ANOVA. There was a significant main effect for warning on all HEXACO traits except for Openness, $F(1, 591) = 2.90$, $p > .10$ (Table 34). In other words, warning lowered test scores for: (1) Honesty Humility, from $M = 3.92$, $SD = .61$ to M = 3.75, SD = .66, $F(1, 591) = 9.88$, $p < .05$, $d = -.26$, (2) Extraversion, from $M = 3.89$, $SD = .75$ to M = 3.66, SD

= .75, $F(1, 591) = 12.97$, $p < .05$, $d = -.31$, (3) Agreeableness, from $M = 3.87$, $SD = .70$ to M = 3.59, SD = .70, $F(1, 591) = 27.63$, $p < .05$, $d = -.43$, and (4) Conscientiousness, from $M = 4.26$, $SD = .55$ to M = 4.16, SD = .52, $F(1, 591) = 5.84$, $p < .05$, $d = -.19$. Warning also raised test scores for Emotionality from $M = 2.74$, $SD = .74$ to M = 2.94, SD = .70, for Emotionality, $F(1, 591) = 10.98$, $p < .05$, $d = .26$. Presumably, this is because in a lower rather than higher score on Emotionality is supposedly more desirable.

Hypothesis 5 predicted that warning would lower motivation to fake. To test Hypothesis 4, an ANOVA indicated that there was only a trend toward lower motivation to fake in the warning ($M = 3.31$, $SD = 1.06$) compared to the no-warning ($M = 3.48$, $SD = 1.08$) condition, $F(1, 584) = 3.63$, $p < .10$, $d = -.16$ (Table 35). In other words, Hypothesis 5 was not supported.

Hypothesis 4 expected that warning would lower faking. An ANOVA showed that the warning had a significant main effect in the expected direction. Specifically, faking was lower in the warning ($M = 1.56$, $SD = .57$) compared to the no-warning ($M = 1.83$, $SD = .75$) condition, $F(1, 576) = 23.10$, $p < .05$, $d = -.41$ (Table 35). Thus, Hypothesis 4 was supported.

The expected correlation between motivation to fake and faking (Hypothesis 6) was also supported. Inspection of the correlation between the two variables in Table 30 showed a small-to-medium correlation in the expected direction between motivation to fake and faking, $r(574) = .23$, $p < .05$.

To test whether the impact of warning on test scores were mediated by motivation to fake, and faking (Hypothesis 7), six separate mediation models were estimated, one for

each HEXACO trait. Specifically, I also conducted a mediation analysis for Openness, even when there was no main effect of warning on the Openness scores. As per Study 2, all mediation models were estimated using the PROCESS macro (model 6) in SPSS (Hayes, 2009). All of the estimates in each mediation model were based on 10,000 bootstrap samples and can be found in Table 36. The parameter estimates for Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openess were reported in Tables 37, 38, 39, 40, 41 and 42 respectively.

To evaluate Hypothesis 7, inspection of the Table 31 indicated that there was no evidence that motivation to fake mediated the effects of warning on test score. This was because all 95% CI estimates for a_1b_1 included zero. Similarly, there was no evidence that the effect of warning on test score was serially mediated by motivation to fake and faking, i.e., all 95% CI estimates for $a_1d_{21}b_2$ included zero. However, there was some evidence that faking mediated the effects of warning on test scores. Specifically, specific indirect effects of faking were found for three of the six HEXACO traits: (1) Honesty-Humility, a_2b_2 = .02 [.01, .04], (2) Conscientiousness, a_2b_2 = .03 [.02, .04], and (3) Openness, a_2b_2 = .02 [.01, .04]. More importantly, the finding that faking mediated the effects of warning on test scores reduces the confidence that a stronger disagreement in the faking measure is simply uninformative. Restated, this should raise the confidence that we have that the faking measure used in Studies 2 and 3 were informative. This implies that we should reject the second account in Study 2, which claimed that the warning was inert.

Nevertheless, when all the three specific indirect (mediation) effects were considered together in Study 3, there was only a statistically significant total indirect

effect for Honesty-Humility, i.e., $a_1b_1 + a_1d_{21}b_2 + a_2b_2 = .02$ [.01, .04], but not for Conscientiousness or Openness. The non-zero total indirect effect for Honesty-Humility was primarily driven by the simple mediation effect via faking (a_2b_2). Further inspection of Table 36 indicated that the overall lack of a statistically significant total indirect effect for Conscientiousness and Openness, despite the observed simple meditation effect via faking, could be explained by examining the other two mediation effects that comprised the total indirect effect. In other words, while the simple mediation effect attributed to faking (a_2b_2) was positive for both Conscientiousness and Openness, both appeared to be partially "offset" by negative simple mediation effects attributed to motivation to fake (a_1b_1). In fact, inspection of the specific indirect effects for the Emotionality, Extraversion and Agreeableness traits indicate that a "balance" between positive and negative specific indirect effects could have contributed towards an overall zero indirect effect. This emphasizes the importance of examining the specific indirect effects rather than relying on the presence of a non-zero total indirect effect, especially when we are interested in probing the specific processes (e.g., motivation and faking) that an intervention (warning) is expected to work via. Overall, there was some evidence that the effect of warning on test score was mediated by motivation and faking, i.e., Hypothesis 7 was partially supported.

To summarize, probing the underlying process model in Study 3 supported the claim that the "prevailing standard" is indeed insufficient to evaluate whether the warning is an effective intervention against faking. Take for example the Openness trait. While there was a zero total effect of warning on test scores (i.e., there was no evidence of a main effect that warning lowered Openness), further inspection of the mediation results

revealed a statistically significant specific indirect effect of faking, as well as a statistically significant direct effect. Clearly, this means that had we only relied on the testing for a main effect of warning on test score, we would have incorrectly concluded that the warning failed to reduce faking for Openness.

In addition, the statistically significant direct effect that remained for Openness even after taking into account the expected mediators suggests that the warning could have caused other systematic but irrelevant effects on test scores. In fact, statistically significant direct effects were observed across all six HEXACO traits. Evidently, the expectation that the effect of warning on test scores is only mediated by motivation to fake and faking is clearly not supported. Overall, the presence of a statistically significant direct effect even after taking into account all of the specific indirect effects suggests that other as of yet unspecified mechanisms were responsible for transmitting the effect of warning on test scores across the six HEXACO traits. Given the evidence so far, it would seem that the warning reduced faking for three of the six traits, but also adversely affected test scores for all six traits.

Cognizant of the evidence that the warning reduced faking, but also led to some unexplained effects, I continued to test Hypothesis 8 to see if the warning contributed to an improvement in the correlation between personality and the pre-manipulation measures of impression management (IM), organizational citizenship behavior (OCB) and counterproductive work behavior (CWB). First, the correlations between the HEXACO trait scores and IM can be found in Table 43. The separate multiple moderated regression (MMR) models that estimated the individual moderation effects for each of the six HEXACO traits and IM were reported in Tables 44, 45, 46, 47, 48, and 49

respectively. Inspection of the MMR results showed that there was only an interaction that trended towards significance between the warning manipulation and IM to predict Honesty-Humility, b_3 = .01, 95% CI [.00, .02], se_{b3} = .08, t(589) = 1.93, p < .10 (Table 44). Specifically, including the interaction term in Model 2 led to a ΔR^2 = .01, ΔF(1, 589) = 3.74, p < .10. This trending interaction suggested that the correlation between Honesty-Humility and IM was stronger in the warning, r(298) = .48 , p < .05 compared to the no-warning condition, r(291) = .38 , p < .05 (Table 43). In contrast, there was no evidence for an interaction effect between warning and IM for the other five HEXACO traits.

The correlations between the HEXACO trait scores and OCB across the warning and no-warning conditions can be found in Table 50. The separate multiple moderated regression (MMR) models that estimated the individual moderation effects for each of the six HEXACO traits were reported in Tables 51, 52, 53, 54, 55 and 56 respectively. Inspection of the MMR results showed that warning did not moderate the strength of association between any of the six HEXACO traits and OCB.

The correlations between the HEXACO trait scores and CWB across the warning and no-warning conditions can be found in Table 57. The separate multiple moderated regression (MMR) models that estimated the individual moderation effects for each of the six HEXACO traits were reported in Tables 58, 59, 60, 61, 62 and 63respectively. Inspection of the MMR results showed that there was only an interaction effect between Warning and Honesty-Humility (Table 58), b_3 = -.70, 95% CI [-1.32, -.07], se_{b3} = .32, t(589) = -2.18, p < .05. Specifically, including the interaction term in Model 2 led to a ΔR^2 = .01, ΔF(1, 589) = 4.77, p < .05. In other words, the warning increased the strength of the correlation between Honesty-Humility and CWB in the warning, r(298) = -.39 , p

< .05 compared to the no-warning condition, $r(291) = -.18$, $p < .05$. Overall, there was minimal support for Hypothesis 8. In other words, the warning only increased the strength of association between HEXACO and another variable for only one out of the eighteen correlations (Honesty-Humility and CWB, Table 57).

Chapter 5: Overall Discussion

Overall, across three studies, I have demonstrated that the effectiveness of warning as an intervention against faking varies depending on the (implicit) definition of faking used. Specifically, previous studies on warning inferred that faking was reduced because warnings reduced test score inflation. While there was some evidence that supports this inference, by actually comparing the "prevailing standard" with an alternative approach that emphasizes test score validity, I have also shown that the "prevailing standard" is insufficient to accurately describe the whole picture. More specifically, in this thesis, I demonstrated with several counter-examples as to why it was insufficient.

Put simply, because faking need not be limited to test score inflation (Burns & Christiansen, 2011), there should be no reason why an effective intervention against faking has to lead to test score deflation only. First, I showed in Study 2 that the warning reduced faking without an impact in test scores. Second, I showed in Study 3 that faking only partially mediated the effect of warning on test scores. Third, I also showed in Study 3 that the remainder of the unexplained indirect effect could not be explained by motivation to fake, even though warning had a main effect on motivation to fake that was trending towards significance.

More importantly, as a result of probing the underlying process model, this thesis also revealed that the warning might not be such a desirable intervention against faking after all. To be precise, while there is evidence that the warning reduced faking, there was also evidence in Study 3 that the warning was a source of systematic but irrelevant

variance on test scores as well. This result has implications for research and practice. First, it highlights a need to at least review past research on warnings and perhaps the broader field of faking and interventions against faking in general with a new lens. To be clear, it would be useful for future research efforts to understand that failing to meet the "prevailing standard" does not mean that the intervention was ineffective at reducing faking. Indeed, as the results in this thesis suggested, future research should instead focus on the examination of the underlying process model with mediation analyses. As such, I cannot emphasize how important it is for the assumption "*that the difference in mean scores is due to a reduction in faking* (Dwight & Donovan, 2003, pg. 2)" to be more fully explored in future studies.

Moreover, given the plausibility of missing/unspecified mediators that could actually "contaminate" the test scores as a result of the warning, it is also perhaps understandable why there was a lack of evidence to support the claim of increased predictive validity by the personality traits in the warning condition. For example, to the extent that the warning was also simultaneously a source of systematic but irrelevant effects in test scores, any overall improvement in predictive validity due to reduced faking could have been "canceled out" because of the introduction of these additional irrelevant effects. In other words, the key implication for practice is that warnings may not be as effective an intervention against faking as we thought. One recommendation is that until we can ensure that we can design a warning such it does not harm test score validity, it may be worth considering a moratorium on the use of warnings against faking in high stakes contexts. Put differently; remember that is vital that in our quest to reduce

faking, we should avoid making things worse, while trying to make things better, i.e., *primum no nocere*.

Lastly, this thesis is not without its limitations. The fundamental limitation in this thesis is that measurement equivalence was only established to be present for respondents that were presumably motivated to fake. In addition, the original baseline model demonstrated less than ideal fit. This result suggests that the measures used in Study 3 may already have psychometric properties that diverged from measures in an "honest-take" condition. However, because there was no third "honest-take" condition, I could not examine whether there was measurement equivalence between test responses in an honest-take versus warning condition to. In other words, this thesis (Study 3) only examined the presence of measurement equivalence between the no-warning versus warning conditions.

As such, future studies should include an honest-take condition, to establish whether the warned test responses can ultimately be compared to honest, not motivated to fake test responses. In other words, establishing measurement equivalence across warning and no-warning conditions was arguably a weak test. A stronger test would be to establishing measurement equivalence across warning condition and an "honest-take" condition.

Appendix 1

!!! IMPORTANT NOTICE !!!

Please complete the personality test honestly. Dishonest responses will NOT increase your chances of getting the top score on the personality test. Indeed, if you respond dishonestly, your results may be invalidated. Therefore, please be truthful in your responses.

Appendix 2

*O*Net Online Report*

O*NET OnLine

Updated 2010

Bright Outlook

Custom Report for:

41-2031.00 - Retail Salespersons

Sell merchandise, such as furniture, motor vehicles, appliances, or apparel to consumers.

Sample of reported job titles: Sales Associate, Sales Consultant, Sales Clerk, Sales Person, Customer Assistant, Clerk, Sales Representative, Design Consultant, Salesman, Bridal Consultant

View report: Summary | Details | Custom

Work Styles

Save Table (XLS/CSV)

Importance	Work Style
87	**Dependability** — Job requires being reliable, responsible, and dependable, and fulfilling obligations.
86	**Cooperation** — Job requires being pleasant with others on the job and displaying a good-natured, cooperative attitude.
86	**Self Control** — Job requires maintaining composure, keeping emotions in check, controlling anger, and avoiding aggressive behavior, even in very difficult situations.
85	**Integrity** — Job requires being honest and ethical.
83	**Attention to Detail** — Job requires being careful about detail and thorough in completing work tasks.
78	**Concern for Others** — Job requires being sensitive to others' needs and feelings and being understanding and helpful on the job.
77	**Initiative** — Job requires a willingness to take on responsibilities and challenges.
75	**Adaptability/Flexibility** — Job requires being open to change (positive or negative) and to considerable variety in the workplace.
75	**Social Orientation** — Job requires preferring to work with others rather than alone, and being personally connected with others on the job.
74	**Achievement/Effort** — Job requires establishing and maintaining personally challenging achievement goals and exerting effort toward mastering tasks.
74	**Independence** — Job requires developing one's own ways of doing things, guiding oneself with little or no supervision, and depending on oneself to get things done.
74	**Stress Tolerance** — Job requires accepting criticism and dealing calmly and effectively with high stress situations.
70	**Persistence** — Job requires persistence in the face of obstacles.

Appendix 3

HEXACO-60 Item Statements

1. I would be quite bored by a visit to an art gallery.* (OP)
2. I plan ahead and organize things, to avoid scrambling at the last minute. (CO)
3. I rarely hold a grudge, even against people who have badly wronged me. (AG)
4. I feel reasonably satisfied with myself overall. (EX)
5. I would feel afraid if I had to travel in bad weather conditions. (EM)
6. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed. (HH)
7. I'm interested in learning about the history and politics of other countries. (OP)
8. I often push myself very hard when trying to achieve a goal. (CO)
9. People sometimes tell me that I am too critical of others.* (AG)
10. I rarely express my opinions in group meetings.* (EX)
11. I sometimes can't help worrying about little things. (EM)
12. If I knew that I could never get caught, I would be willing to steal a million dollars.* (HH)
13. I would enjoy creating a work of art, such as a novel, a song, or a painting. (OP)
14. When working on something, I don't pay much attention to small details.* (CO)
15. People sometimes tell me that I'm too stubborn.* (AG)
16. I prefer jobs that involve active social interaction to those that involve working alone. (EX)
17. When I suffer from a painful experience, I need someone to make me feel comfortable. (EM)

18. Having a lot of money is not especially important to me. (HH)
19. I think that paying attention to radical ideas is a waste of time.* (OP)
20. I make decisions based on the feeling of the moment rather than on careful thought.* (CO)
21. People think of me as someone who has a quick temper.* (AG)
22. On most days, I feel cheerful and optimistic.(EX)
23. I feel like crying when I see other people crying. (EM)
24. I think that I am entitled to more respect than the average person is.* (HH)
25. If I had the opportunity, I would like to attend a classical music concert. (OP)
26. When working, I sometimes have difficulties due to being disorganized.* (CO)
27. My attitude toward people who have treated me badly is “forgive and forget”. (AG)
28. I feel that I am an unpopular person.* (EX)
29. When it comes to physical danger, I am very fearful. (EM)
30. If I want something from someone, I will laugh at that person's worst jokes.* (HH)
31. I’ve never really enjoyed looking through an encyclopedia.* (OP)
32. I do only the minimum amount of work needed to get by.* (CO)
33. I tend to be lenient in judging other people. (AG)
34. In social situations, I’m usually the one who makes the first move. (EX)
35. I worry a lot less than most people do.* (EM)
36. I would never accept a bribe, even if it were very large. (HH)
37. People have often told me that I have a good imagination. (OP)

38. I always try to be accurate in my work, even at the expense of time. (CO)
39. I am usually quite flexible in my opinions when people disagree with me. (AG)
40. The first thing that I always do in a new place is to make friends. (EX)
41. I can handle difficult situations without needing emotional support from anyone else.* (EM)
42. I would get a lot of pleasure from owning expensive luxury goods.* (HH)
43. I like people who have unconventional views. (OP)
44. I make a lot of mistakes because I don't think before I act.* (CO)
45. Most people tend to get angry more quickly than I do. (AG)
46. Most people are more upbeat and dynamic than I generally am.* (EX)
47. I feel strong emotions when someone close to me is going away for a long time. (EM)
48. I want people to know that I am an important person of high status.* (HH)
49. I don't think of myself as the artistic or creative type.* (OP)
50. People often call me a perfectionist. (CO)
51. Even when people make a lot of mistakes, I rarely say anything negative. (AG)
52. I sometimes feel that I am a worthless person.* (EX)
53. Even in an emergency I wouldn't feel like panicking.* (EM)
54. I wouldn't pretend to like someone just to get that person to do favors for me. (HH)
55. I find it boring to discuss philosophy.* (OP)
56. I prefer to do whatever comes to mind, rather than stick to a plan.* (CO)

57. When people tell me that I'm wrong, my first reaction is to argue with them.* (AG)
58. When I'm in a group of people, I'm often the one who speaks on behalf of the group. (EX)
59. I remain unemotional even in situations where most people get very sentimental.* (EM)
60. I'd be tempted to use counterfeit money, if I were sure I could get away with it.* (HH)

*Note. * = reverse-coded; HH = Honesty-Humility, EM = Emotionality, EX = Extraversion, AG = Agreeableness, CO = Conscientiousness, OP = Openness to Experience.*

Appendix 4

Study 1. Faking Behavior Item Statements

1. I overemphasized or exaggerated my positive attributes.
2. I tended to de-emphasize or “play-down” what some might consider my negative attributes.
3. I exaggerated qualities or characteristics of myself such as dependability and reliability.
4. I gave responses that were completely false or made-up.
5. I exaggerated my responses to make myself look better than I actually am.
6. I exaggerated less than 10% of the information I provided.
7. I exaggerated about a quarter (25%) of the information I provided.
8. I exaggerated about half (50%) of the information I provided.
9. I exaggerated about three quarters (75% or more) of the information I provided.
10. I made up less than 10% of the information I provided.
11. I made up about a quarter (25%) of the information I provided.
12. I made up about half (50%) of the information I provided.
13. I made up about three quarters (75%) of the information I provided.
14. I outright fabricated or made-up information all the time.

Appendix 5

mini-IPIP Item Statements

1. Am the life of the party.(E)
2. Sympathize with others' feelings. (A)
3. Get chores done right away. (C)
4. Have frequent mood swings. (N)
5. Have a vivid imagination. (I)
6. Don't talk a lot.* (E)
7. Am not interested in other people's problems.* (A)
8. Often forget to put things back in their proper place.* (C)
9. Am relaxed most of the time.* (N)
10. Am not interested in abstract ideas.* (I)
11. Talk to a lot of different people at parties. (E)
12. Feel others' emotions. (A)
13. Like order. (C)
14. Get upset easily. (N)
15. Have difficulty understanding abstract ideas.* (I)
16. Keep in the background.* (E)
17. Am not really interested in others.* (A)
18. Make a mess of things.* (C)
19. Seldom feel blue.* (N)
20. Do not have a good imagination.* (I)

Note. * = *reverse-coded; E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, I = Intellect/Imagination.*

Appendix 6

Study 2. Faking Behavior Item Statements

1. I am as agreeable as I reported myself to be.*
2. I am actually less sociable than I presented myself to be.
3. I tried to portray myself as more friendly than I really am.
4. I presented an accurate image of myself.*
5. I am as conscientious as I reported myself to be.*
6. I am actually more emotional than I presented myself to be.
7. I tried to portray myself as more hardworking than I really am.
8. I was completely honest in my responses.*

*Note. * = reverse-coded.*

Appendix 7

Organizational Citizenship Behavior Checklist (OCB-C) item statements

1. Took time to advise, coach, or mentor a co-worker.
2. Helped a co-worker learn new skills or shared job knowledge.
3. Helped new employees get oriented to the job.
4. Lent a compassionate ear when someone had a work problem.
5. Offered suggestions to improve how work is done.
6. Helped a co-worker who had too much to do.
7. Volunteered for extra work assignments.
8. Worked weekends or other days off to complete a project or task.
9. Volunteered to attend meetings or work committees on your own time.
10. Gave up meal and other breaks to complete work.

Appendix 8

Counter-productive Work Behavior Checklist (CWB-C) item statements

1. Purposely wasted your employer's materials/supplies.
2. Complained about insignificant things at work.
3. Told people outside the job what a lousy place you work for.
4. Came to work late without permission.
5. Stayed home from work and said you were sick when you weren't.
6. Insulted someone about their job performance.
7. Made fun for someone's personal life.
8. Ignored someone at work.
9. Started an argument with someone at work.
10. Insulted or made fun of someone at work.

Appendix 9

Impression Management (IM) Subscale of Balanced Inventory of Desirable Responding (BIDR) Item Statements

1. I sometimes tell lies if I have to.*
2. I never cover up my mistakes.
3. There have occasions where I have taken advantage of someone.*
4. I never swear.
5. I sometimes try to get even rather than forgive and forget.*
6. I always obey laws, even if I'm unlikely to get caught.
7. I have said something bad about a friend behind his/her back.*
8. When I hear people talking privately, I avoid listening.
9. I have received too much change from a salesperson without telling him or her.*
10. I always declare everything at customs.
11. When I was young, I sometimes stole things.*
12. I have never dropped litter on the street.
13. I sometimes drive faster than the speed limit.*
14. I have never read sexy books or magazines.
15. I have done things that I don't tell people about.*
16. I never take things that don't belong to me.
17. I have taken sick leave from work or school even though I wasn't really sick.*
18. I have never damaged a library book or store merchandise without reporting it.
19. I have some pretty awful habits.*
20. I don't gossip about other people's business.

Note. * = *reverse-coded.*

Appendix 10

Study 3. R Code for Tests of Measurement Equivalence

```
# specify the CFA model to be tested

s3.model <-

  'HH =~ h6 + h30r + h54 + h12r + h36 + h60r + h18 + h42r + h24r + h48r

  EM =~ h5 + h29 + h53r + h11 + h35r + h17 + h41r + h23 + h47 + h59r

  EX =~ h4 + h28r + h52r + h10r + h34 + h58 + h16 + h40 + h22 + h46r

  AG =~ h3 + h27 + h9r + h33 + h51 + h15r + h39 + h57r + h21r + h45

  CO =~ h2 + h26r + h8 + h32r + h14r + h38 + h50 + h20r + h44r + h56r

  OP =~ h1r + h25 + h7 + h31r + h13+ h37 + h49r + h19r + h43 + h55r

  MOT =~ f10 + f11 + f12r

  FAK =~ f2 + f3 + f4 + f5r + f6r + f7r + f8 + f9r'


# model 0: baseline model

model0 <- cfa (s3.model, data=s3)

summary (model0, fit.measures=TRUE)


# model 1: configural equivalence (same form) model
```

```
model1 <- cfa(s3.model, data=s3, group="warning")

summary (model1, fit.measures=TRUE)


# model 2: metric equivalence (same item-factor loading) model

model2 <- cfa(s3.model, data=s3, group="warning", group.equal=c("loadings"))

summary (model2, fit.measures=TRUE)


# compare model 1 vs. model 2

anova(model1, model2)


# model 3: scalar equivalence (same item intercept) model

model3 <- cfa(s3.model, data=s3, group="warning", group.equal=c("loadings",
"intercepts"))

summary (model3, fit.measures=TRUE)


# compare model 2 vs. model 3

anova(model2, model3)
```

Table 1

Study 1 - Correlation Table (N = 190)

	α	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.
1. Warning		0.00	1.00							
2. Incentive		0.05	1.00	-.02						
3. Honesty-Humility	.79	3.56	0.71	-.09	.05					
4. Emotionality	.75	2.98	0.67	-.03	.03	-.14*				
5. Extraversion	.88	3.46	0.81	-.14^{+}	.05	.31*	-.47*			
6. Agreeableness	.86	3.46	0.79	-.22*	-.02	.42*	-.38*	.64*		
7. Conscientiousness	.82	3.92	0.65	-.02	-.02	.49*	-.30*	.55*	.54*	
8. Openness	.84	3.70	0.73	-.08	-.01	.25*	-.17*	.40*	.39*	.47*

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 2

Study 1 - Faking Behavior Item Responses (N = 190)

Item Statement	Incentive And Warning	Incentive	Warning	No Incentive Nor Warning
1	0.14	0.37	0.09	0.18
2	0.10	-0.02	0.39	0.36
3	0.14	0.18	0.17	-0.09
4	-0.27	0.02	0.09	0.14
5	0.18	0.02	-0.04	0.05
6	0.55	0.49	0.61	0.50
7	-0.27	0.02	0.09	-0.14
8	-0.14	0.06	0.13	-0.36
9	-0.22	-0.25	0.00	-0.23
10	0.35	0.57	0.57	0.32
11	-0.22	-0.02	-0.17	0.00
12	-0.10	-0.02	-0.09	-0.18
13	-0.18	-0.18	-0.04	-0.41
14	-0.06	-0.14	-0.04	-0.14
n	49	51	46	44
Average Proportion	.10	.12	.15	.11
z	2.27	2.55	2.26	2.26
p	.02	.01	.02	.02

Note. Values indicate proportion of "True" responses to item statements. At the item level, proportion was calculated by using the following formula: ([sum of "True"

responses] - [$n_{condition}$ / 2]) / ($n_{condition}$ / 2). For example, for item 1 in condition 1: sum of "True" responses = 28, i.e., proportion = (28 – [49/2]) / (49/2) = 3.5 / 24.5 = .14. At the condition level, negative item proportions were treated as zero when calculating the average proportion across 14 items.

Table 3

Study 1. One-Way ANOVA of HEXACO Traits by Incentive

	No Incentive (n = 90)		Incentive (n = 100)				
	M	*SD*	*M*	*SD*	*d*	df_1, df_2	*F*
Honesty-Humility	3.52	.71	3.60	.72	.11	1, 188	.46
Emotionality	2.96	.68	3.00	.67	.06	1, 188	.12
Extraversion	3.42	.76	3.50	.86	.10	1, 188	.38
Agreeableness	3.48	.73	3.44	.84	-.05	1, 188	.07
Conscientiousness	3.90	.63	3.93	.67	.05	1, 188	.08
Openness	3.70	.70	3.69	.76	-.01	1, 188	.01

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 4

Study 1. One-Way ANOVA of HEXACO Traits by Warning

	No Warning (n = 95)		Warning (n = 95)				
	M	*SD*	*M*	*SD*	*d*	df_1, df_2	*F*
Honesty-Humility	3.62	.71	3.50	.71	-.17	1, 188	1.41
Emotionality	3.00	.68	2.96	.66	-.06	1, 188	.19
Extraversion	3.57	.78	3.35	.84	-.32	1, 188	3.52^{+}
Agreeableness	3.63	.78	3.29	.76	-.44	1, 188	9.21*
Conscientiousness	3.93	.68	3.90	.63	-.05	1, 188	.08
Openness	3.76	.73	3.64	.73	-.16	1, 188	1.11

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 5

Study 1. Completely Randomized Factorial (CRF-2, 2) Results for Incentive by Warning Interaction for HEXACO Traits

	No Incentive					Incentive				
	No Warning (n = 44)		Warning (n = 46)			No Warning (n = 51)		Warning (n = 49)		
	M	*SD*	*M*	*SD*	*d*[1]	*M*	*SD*	*M*	*SD*	*d*[1]
Honesty-Humility	3.60	.73	3.46	.69	-.20	3.64	.69	3.54	.74	-.14
Emotionality	2.92	.69	3.01	.67	.13	3.08	.68	2.92	.65	-.24
Extraversion	3.52	.77	3.33	.75	-.25	3.62	.78	3.37	.92	-.29
Agreeableness	3.67	.76	3.29	.66	-.53	3.59	.80	3.29	.85	-.36
Conscientiousness	3.94	.64	3.87	.63	-.15	3.92	.71	3.93	.63	.01
Openness	3.67	.74	3.74	.67	.10	3.83	.73	3.55	.78	-.37

Note. [1] = *d value is for the simple main effect of warning;* [+] = $p < .10$, * = $p < .05$.

Table 6

Study 1. Completely Randomized Factorial (CRF-2, 2) Source Tables for Incentive by Warning Interaction for HEXACO Traits

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Honesty-Humility	Incentive	1	.22	.22	.42	.002
	Warning	1	.71	.71	1.39	.007
	Interaction	1	.02	.02	.03	.000
	Error	186	95.06	.51		
	Total	189	96.00			
Emotionality	Incentive	1	.05	.05	.11	.001
	Warning	1	.06	.06	.13	.001
	Interaction	1	.74	.74	1.64	.009
	Error	186	84.12	.45		
	Total	189	84.99			
Extraversion	Incentive	1	.22	.22	.34	.002
	Warning	1	2.23	2.23	3.38^{+}	.018
	Interaction	1	.05	.05	.08	.000
	Error	186	122.46	.66		
	Total	189	125.03			
Agreeableness	Incentive	1	.07	.07	.12	.001
	Warning	1	5.51	5.51	9.23^{*}	.047
	Interaction	1	.06	.06	.10	.001
	Error	186	110.99	.60		
	Total	189	116.56			

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Conscientiousness	Incentive	1	.03	.03	.08	.000
	Warning	1	.04	.04	.09	.000
	Interaction	1	.09	.09	.21	.001
	Error	186	79.98	.43		
	Total	189	80.14			
Openness	Incentive	1	.01	.01	.01	.000
	Warning	1	.50	.50	.93	.005
	Interaction	1	1.48	1.48	2.79^{+}	.015
	Error	186	98.68	.53		
	Total	189	100.76			

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 7.

Study 2. Correlation Table (N varies from 106 to 115)

	α	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Warning		-.03	1.00												
2. Incentive		-.03	1.00	.04											
3. Extraversion[1]	.80	2.80	1.00	.10	-.10										
4. Agreeableness[1]	.77	3.81	.83	.05	-.07	.28*									
5. Conscientiousness[1]	.80	3.72	.93	.04	-.04	-.00	.04								
6. Neuroticism[1]	.78	2.56	.98	-.17^{+}	-.01	-.35*	-.04	-.40*							
7. Intellect[1]	.78	3.71	.92	-.01	.07	.18^{+}	.22*	.02	-.17^{+}						
8. Extraversion[2]	.90	3.73	.82	.01	.02	.47*	.02	.12	-.33*	.08					
9. Agreeableness[2]	.82	3.71	.71	.04	.04	.23*	.17^{+}	.17^{+}	-.30*	.06	.70*				
10. Conscientiousness[2]	.80	4.12	.62	.02	.09	.09	.10	.34*	-.25*	.16	.63*	.48*			
11. Emotionality[2]	.81	2.80	.77	-.09	-.12	-.22*	.40*	-.14	.38*	-.07	-.54*	-.36*	-.37*		
12. Motivation to fake	.73	3.92	.81	-.02	.00	.06	.09	.07	-.13	-.11	.36*	.25*	.33*	-.08	
13. Faking behavior	.77	1.91	.77	-.23*	-.08	-.17^{+}	-.20*	-.40*	.45*	-.07	.00	-.00	-.13	-.04	-.04

Note. [1] = mini-IPIP (Administered Before Manipulation), [2] = HEXACO (Administered After Manipulation); $^{+}$ = $p < .10$, * = $p < .05$.

Table 8

Study 2. One-Way ANOVA of mini-IPIP Traits by Incentive

	No Incentive			Incentive					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Extraversion	2.88	1.05	56	2.69	.96	55	-.19	1, 109	1.02
Agreeableness	3.86	.78	57	3.75	.88	55	-.13	1, 110	.53
Conscientiousness	3.77	.84	56	3.70	1.01	55	-.12	1, 109	.17
Neuroticism	2.57	.96	57	2.56	1.00	54	-.01	1, 109	.01
Intellect	3.65	.95	57	3.78	.90	54	.14	1, 109	.50

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 9

Study 2. One-Way ANOVA of mini-IPIP Traits by Warning

	No Warning			Warning					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Extraversion	2.69	.95	58	2.89	1.07	53	.20	1, 109	1.07
Agreeableness	3.76	.90	58	3.85	.76	54	.11	1, 110	.29
Conscientiousness	3.70	.95	57	3.77	.91	54	.08	1, 109	.16
Neuroticism	2.72	.96	57	2.39	.97	54	-.34	1, 109	3.24^{+}
Intellect	3.72	.84	57	3.71	1.01	54	-.01	1, 109	.00

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 10

Study 2. Completely Randomized Factorial (CRF-2, 2) Results for Incentive by Warning Interaction for mini-IPIP Traits

	No Incentive					Incentive				
	No Warning		Warning			No Warning		Warning		
	M (SD)	*n*	*M (SD)*	*n*	*d*[1]	*M (SD)*	*n*	*M (SD)*	*n*	*d*[1]
Extraversion	2.72 (1.05)	31	3.09 (1.04)	25	.35	2.67 (.83)	27	2.71 (1.08)	28	.04
Agreeableness	3.67 (.81)	31	4.07 (.71)	26	.52	3.85 (1.00)	27	3.64 (.76)	28	-.24
Conscientiousness	3.84 (.81)	30	3.68 (.89)	26	-.19	3.53 (1.07)	27	3.85 (.94)	28	.32
Neuroticism	2.71 (.95)	31	2.40 (.96)	26	-.32	2.74 (1.00)	26	2.38 (.99)	28	-.36
Intellect	3.61 (.89)	31	3.70 (1.04)	26	.09	3.85 (.77)	26	3.71 (1.01)	28	-.16

Note. [1] *= d value is for the simple main effect of warning;* $^{+}$ = $p < .10$, * = $p < .05$.

Table 11

Study 2. Completely Randomized Factorial (CRF-2, 2) Source Tables for Incentive by Warning Interaction for mini-IPIP Traits

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Extraversion	Incentive	1	1.26	1.26	1.24	.01
	Warning	1	1.22	1.22	1.20	.01
	Interaction	1	.73	.73	.72	.01
	Error	107	108.54	1.01		
	Total	110	111.53			
Agreeableness	Incentive	1	.46	.46	.68	.01
	Warning	1	.21	.21	.31	.00
	Interaction	1	2.43	2.43	3.56^{+}	.03
	Error	108	73.78	.68		
	Total	111	76.80			
Conscientiousness	Incentive	1	.13	.13	.16	.00
	Warning	1	.16	.16	.19	.00
	Interaction	1	1.53	1.53	1.77	.02
	Error	107	92.51	.87		
	Total	110	94.34			
Neuroticism	Incentive	1	.00	.00	.00	.00
	Warning	1	3.03	3.03	3.19^{+}	.03
	Interaction	1	.02	.02	.02	.00
	Error	107	101.64	.95		
	Total	110	104.68			

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Intellect	Incentive	1	.42	.42	.48	.00
	Warning	1	.01	.01	.02	.00
	Interaction	1	.34	.34	.39	.00
	Error	107	92.96	.87		
	Total	110	93.73			

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 12

Study 2. One-Way ANOVA of HEXACO Traits by Incentive

	No Incentive			Incentive					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Extraversion	3.71	.81	57	3.74	.83	54	.04	1, 109	.03
Agreeableness	3.68	.67	57	3.73	.75	54	.07	1, 109	.15
Conscientiousness	4.06	.62	57	4.18	.61	54	.20	1, 109	.97
Emotionality	2.88	.66	57	2.71	.76	54	-.24	1, 109	1.61

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 13

Study 2. One-Way ANOVA of HEXACO Traits by Warning

	No Warning			Warning					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Extraversion	3.72	.82	56	3.73	.83	55	.01	1, 109	.01
Agreeableness	3.68	.75	56	3.74	.67	55	.08	1, 109	.21
Conscientiousness	4.10	.64	56	4.13	.60	55	.05	1, 109	.05
Emotionality	2.86	.71	56	2.73	.72	55	-.18	1, 109	.38

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 14

Study 2. Completely Randomized Factorial (CRF-2, 2) Results for Incentive by Warning Interaction for HEXACO Traits

	No Incentive					Incentive				
	No Warning		Warning			No Warning		Warning		
	M (SD)	*N*	*M (SD)*	*n*	*d*[1]	*M (SD)*	*n*	*M (SD)*	*n*	*d*[1]
Extraversion	3.68 (.88)	30	3.75 (.76)	27	.08	3.77 (.76)	26	3.72 (.90)	28	-.06
Agreeableness	3.67 (.74)	30	3.69 (.61)	27	.03	3.68 (.77)	26	3.78 (.74)	28	.13
Conscientiousness	4.02 (.66)	30	4.11 (.59)	27	.14	4.20 (.61)	26	4.15 (.62)	28	-.08
Emotionality	2.88 (.60)	30	2.88 (.73)	27	.00	2.83 (.82)	26	2.70 (.70)	28	-.17

Note. [1] *= d value is for the simple main effect of warning;* [+] *= p* < .10, * = *p* < .05.

Table 15

Study 2. Completely Randomized Factorial (CRF-2, 2) Source Tables for Incentive by Warning Interaction for HEXACO Traits

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Extraversion	Incentive	1	.20	.20	.03	.00
	Warning	1	.00	.00	.01	.00
	Interaction	1	.12	.12	.17	.00
	Error	107	73.68	.69		
	Total	110	73.82			
Agreeableness	Incentive	1	.07	.07	.13	.00
	Warning	1	.10	.10	.19	.00
	Interaction	1	.05	.05	.09	.00
	Error	107	55.18	.52		
	Total	110	55.40			
Conscientiousness	Incentive	1	.36	.36	.93	.01
	Warning	1	.01	.01	.02	.00
	Interaction	1	.14	.14	.37	.00
	Error	107	41.21	.39		
	Total	110	41.73			
Emotionality	Incentive	1	.77	.77	1.52	.01
	Warning	1	.38	.38	.74	.01
	Interaction	1	.36	.36	.71	.01
	Error	107	54.54	.51		
	Total	110	56.08			

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 16

Study 2. One-Way ANOVA of Motivation to Fake and Faking Behavior by Incentive

	No Incentive			Incentive					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Motivation to Fake	3.92	.79	56	3.92	.83	52	.00	1, 106	.00
Faking	1.98	.65	55	1.85	.87	52	-.17	1, 105	.75

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 17

Study 2. One-Way ANOVA of Motivation to Fake and Faking Behavior by Warning

	No Warning			Warning					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Motivation to Fake	3.93	.78	54	3.91	.84	54	-.02	1, 106	.03
Faking	2.08	.82	57	1.74	.66	52	-.45	1, 105	5.78*

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 18

Study 2. Completely Randomized Factorial (CRF-2, 2) Results for Incentive by Warning Interaction for Motivation to Fake and Faking Behavior

	No Incentive					Incentive				
	No Warning		Warning			No Warning		Warning		
	M (SD)	*N*	*M (SD)*	*n*	*d*[1]	*M (SD)*	*n*	*M (SD)*	*n*	*d*[1]
Motivation to Fake	3.91 (.73)	29	3.93 (.87)	27	.03	3.96 (.85)	25	3.89 (.82)	27	-.08
Faking	2.11 (.63)	29	1.83 (.66)	26	-.43	2.05 (1.01)	26	1.64 (.66)	26	-.48

Note. [1] = *d value is for the simple main effect of warning;* [+] = $p < .10$, * = $p < .05$.

Table 19

Study 2. Completely Randomized Factorial (CRF-2, 2) Source Tables for Incentive by Warning Interaction for Motivation to Fake and Faking Behavior

	Source	*df*	*SS*	*MS*	*F*	Partial η^2
Motivation to Fake	Incentive	1	.00	.00	.00	.00
	Warning	1	.02	.02	.03	.00
	Interaction	1	.05	.05	.08	.00
	Error	104	69.46	.69		
	Total	107	69.53			
Faking Behavior	Incentive	1	.39	.39	.69	.01
	Warning	1	3.21	3.21	5.65*	.05
	Interaction	1	.10	.10	.18	.00
	Error	103	58.54	.57		
	Total	106	62.27			

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 20

Study 2. Coefficient Estimates and 95% Confidence Intervals for Mediation Models

Y	Specific indirect effect			Total indirect effect	Direct effect	Total effect
	$X \rightarrow M_1 \rightarrow Y$	$X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$	$X \rightarrow M_2 \rightarrow Y$			
	a_1b_1	$a_1d_{21}b_2$	a_2b_2		c^1	$a_1b_1 + a_1 d_{21}b_2 + a_2b_2 + c^1$
Extraversion	-.01 (-.07, .04)	.00 (.00, .00)	-.01 (-.06, .03)	-.02 (-.09, .05)	.02 (-.13, .18)	.01 (-.15, .17)
Agreeableness	-.01 (-.05, .03)	.00 (.00, .00)	-.01 (-.05, .03)	-.01 (-.07, .04)	.05 (-.08, .19)	.04 (-.09, .18)
Conscientiousness	-.01 (-.05, .03)	.00 (-.01, .00)	.01 (-.01, .05)	.00 (-.04, .05)	.01 (-.10, .12)	.02 (-.09, .13)
Emotionality	.00 (-.01, .03)	.00 (.00, .00)	.01 (-.02, .06)	.01 (-.02, .06)	-.08 (-.22, .07)	-.06 (-.20, .07)

Note. X =Warning, M_1 = Motivation to fake, M_2 = Faking Behavior, $^+ = p < .10$, $* = p < .05$.

Table 21

Study 2. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Extraversion

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Extraversion)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*P*
X (Warning)	a_1	-.03	.08	> .10	a_2	-.17	.07	*	c^1	.02	.08	> .10
M_1 (Motivation to Fake)					d_{21}	-.05	.09	> .10	b_1	.35	.09	*
M_2 (Faking)									b_2	.04	.10	> .10
Constant	i_{M1}	3.91	.08	*	i_{M2}	2.08	.36	*	i_Y	2.28	.43	*
		R^2 = .00				R^2 = .05				R^2 = .12		
		$F(1, 104) = .12$, $p > .10$				$F(2, 103) = 2.71$, $p < .10$				$F(3, 102) = 4.67$, $p < .05$		

Note. $^+ = p < .10$, $* = p < .05$.

Table 22

Study 2. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Agreeableness

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Agreeableness)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.03	.08	> .10	a_2	-.17	.07	*	c^1	.05	.07	> .10
M_1 (Motivation to Fake)					d_{21}	-.05	.09	> .10	b_1	.22	.08	*
M_2 (Faking)									b_2	.04	.09	> .10
Constant	i_{M1}	3.91	.08	*	i_{M2}	2.08	.36	*	i_Y	2.79	.38	*
		R^2 = .00				R^2 = .05				R^2 = .07		
		$F(1, 104) = .12$, $p > .10$				$F(2, 103) = 2.71$, $p < .10$				$F(3, 102) = 2.55$, $p < .10$		

Note. $^+ = p < .10$, $* = p < .05$.

Table 23

Study 2. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Conscientiousness

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Conscientiousness)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.03	.08	> .10	a_2	-.17	.07	*	c^1	.01	.06	> .10
M_1 (Motivation to Fake)					d_{21}	-.05	.09	> .10	b_1	.23	.07	*
M_2 (Faking)									b_2	-.06	.07	> .10
Constant	i_{M1}	3.91	.08	*	i_{M2}	2.08	.36	*	i_Y	3.34	.31	*
		R^2 = .00				R^2 = .05				R^2 = .11		
		$F(1, 104) = .12$, $p > .10$				$F(2, 103) = 2.71$, $p < .10$				$F(3, 102) = 4.41$, $p < .05$		

Note. $^+ = p < .10$, $* = p < .05$.

Table 24

Study 2. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Emotionality

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Emotionality)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.03	.08	> .10	a_2	-.17	.07	*	c^1	-.08	.07	> .10
M_1 (Motivation to Fake)					d_{21}	-.05	.09	> .10	b_1	-.06	.09	> .10
M_2 (Faking)									b_2	-.06	.09	> .10
Constant	i_{M1}	3.91	.08	*	i_{M2}	2.08	.36	*	i_Y	3.16	.40	*
		R^2 = .00				R^2 = .05				R^2 = .02		
		$F(1, 104) = .12$, $p > .10$				$F(2, 103) = 2.71$, $p < .10$				$F(3, 102) = .57$, $p > .10$		

Note. $^+ = p < .10$, $* = p < .05$.

Table 25

Study 2. Correlations between Pre- and Post-Manipulation Personality across Experimental Conditions

			Warning Absent *r* (df)			Warning Present *r* (df)
	Control *r* (df)	Incentive Only *r* (df)		Warning Only *r* (df)	Incentive and Warning *r* (df)	
Extraversion	.54* (28)	.31 (24)	.46* (54)	.60* (23)	.43* (26)	.49* (51)
Agreeableness	.15 (28)	.20 (24)	.18 (54)	.42* (24)	.02 (26)	.16 (52)
Conscientiousness	.40* (27)	.39* (24)	.36* (53)	.14 (24)	.47* (26)	.32* (52)
Neuroticism/Emotionality	.12 (28)	.33 (23)	.24^{+} (53)	.62* (24)	.44* (26)	.52* (52)

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 26

Study 2. Regression Results for Interaction among Pre-Manipulation Extraversion, Incentive and Warning to Predict Post-Manipulation Extraversion

		Model 1			Model 2			Model 3		
		R^2 = .23, *MSE* = .54			R^2 = .23, *MSE* = .55			R^2 = .23, *MSE* = .56		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	2.60	.21	12.17*	2.61	.23	11.53*	2.62	.23	11.53*
Pre-Manipulation Extraversion (*X*)	b_1	.40	.07	5.58*	.39	.08	5.09*	.39	.08	5.09*
Incentive (M_1)	b_2	.06	.07	.83	.23	.22	1.00	.23	.23	1.01
Warning (M_2)	b_3	-.03	.07	-.46	-.07	.22	-.31	-.08	.23	-.33
$X \times M_1$	b_4				-.06	.08	-.78	-.06	.08	-.80
$X \times M_2$	b_5				.01	.08	.15	.01	.08	.18
$M_1 \times M_2$	b_6				.01	.07	.20	-.03	.23	-.12
$X \times M_1 \times M_2$	b_7							.01	.08	.19

Note. $^+ = p < .10$, * = $p < .05$.

Table 27

Study 2. Regression Results for Interaction among Pre-Manipulation Agreeableness, Incentive and Warning to Predict Post-Manipulation Agreeablness

		Model 1			Model 2			Model 3		
		R^2 = .03, *MSE* = .49			R^2 = .04, *MSE* = .50			R^2 = .05, *MSE* = .50		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.17	.32	10.02*	3.10	.34	9.21*	3.09	.34	9.17*
Pre-Manipulation Agreeableness (*X*)	b_1	.15	.08	1.80^+	.16	.09	1.86^+	.16	.09	1.86^+
Incentive (M_1)	b_2	.02	.07	.34	.25	.33	.76	.32	.34	.94
Warning (M_2)	b_3	.04	.07	.56	.02	.33	.06	-.03	.34	-.08
$X \times M_1$	b_4				-.06	.08	-.70	-.08	.09	-.88
$X \times M_2$	b_5				.00	.09	.03	.01	.09	.15
$M_1 \times M_2$	b_6				.03	.07	.45	.34	.34	1.01
$X \times M_1 \times M_2$	b_7							-.08	.09	-.94

Note. $^+ = p < .10$, * = $p < .05$.

Table 28

Study 2. Regression Results for Interaction among Pre-Manipulation Conscientiousness, Incentive and Warning to Predict Post-Manipulation Conscientiousness

		Model 1 R^2 = .13, *MSE* = .34			Model 2 R^2 = .14, *MSE* = .34			Model 3 R^2 = .15, *MSE* = .34		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.27	.23	14.19*	3.27	.24	13.77*	3.25	.24	13.64*
Pre-Manipulation Conscientiousness (*X*)	b_1	.23	.06	3.80*	.23	.06	3.69*	.23	.06	3.78*
Incentive (M_1)	b_2	.06	.06	1.13	-.04	.24	-.17	-.05	.24	-.21
Warning (M_2)	b_3	-.01	.06	-.24	.07	.24	.28	.12	.24	.48
$X \times M_1$	b_4				.03	.06	.45	.03	.06	.49
$X \times M_2$	b_5				-.02	.06	-.36	-.04	.06	-.57
$M_1 \times M_2$	b_6				-.05	.06	-.91	-.35	.24	-1.48
$X \times M_1 \times M_2$	b_7							.08	.06	1.30

Note. $^+ = p < .10$, $* = p < .05$

Table 29

Study 2. Regression Results for Interaction among Pre-Manipulation Neuroticism, Incentive and Warning to Predict Post-Manipulation Emotionality

		Model 1			Model 2			Model 3		
		R^2 = .16, *MSE* = .45			R^2 = .19, *MSE* = .45			R^2 = .20, *MSE* = .44		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	2.08	.18	11.42*	2.10	.18	11.52*	2.08	.18	11.48*
Pre-Manipulation Neuroticism (*X*)	b_1	.28	.07	4.21*	.28	.07	4.22*	.29	.07	4.30*
Incentive (M_1)	b_2	-.08	.06	-1.31	-.10	.18	-.56	-.11	.18	-.63
Warning (M_2)	b_3	-.01	.07	-.20	-.29	.18	-.40	-.29	.18	-1.61
$X \times M_1$	b_4				.01	.07	.12	.01	.07	.09
$X \times M_2$	b_5				.11	.07	1.62	.11	.07	1.65
$M_1 \times M_2$	b_6				-.05	.07	-.78	.19	.18	1.03
$X \times M_1 \times M_2$	b_7							-.09	.07	-1.40

Note. $^+ = p < .10$, $* = p < .05$

Table 30

Study 3. Correlation Table (N varies from 576 to 620)

	α	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. OCB-C	.89	29.65	7.61											
2. CWB-C	.84	11.61	5.28	.15*										
3. IM	.84	8.05	4.59	-.50*	-.50*									
4. Warning		.01	1.00	-.02	-.02	.02								
5. Honesty-Humility	.76	3.84	.64	.03	-.29*	.43*	-.13*							
6. Emotionality	.82	2.84	.73	-.14*	.11*	-.12*	.14*	-.26*						
7. Extraversion	.89	3.77	.76	.29*	-.20*	.12*	-.15*	.19*	-.46*					
8. Agreeableness	.85	3.74	.70	.10*	.-35*	.31*	-.21*	.47*	-.39*	.49*				
9. Conscientiousness	.84	4.21	.54	.21*	-.28*	.28*	-.10*	.40*	-.33*	.50*	.40*			
10. Openness	.83	4.01	.64	.18*	-.18*	.13*	-.07+	.22*	-.28*	.38*	.37*	.36*		
11. Motivation to Fake	.79	3.39	1.07	.10*	-.03	-.03	-.08+	-.09*	-.12*	.33*	.26*	.22*	.15*	
12. Faking behavior	.87	1.69	.68	-.10*	.25*	-.25*	-.20*	-.15*	.00	-.02	.00	-.21*	-.13*	.23*

Note. OCB-C= Organizational Citizenship Behavior – Checklist; CWB-C = Counterproductive Work Behavior – Checklist; IM = Impression Management; + = $p < .10$, * = $p < .05$.

Table 31

Study 3. One-Way ANOVA of Pre-Manipulation Measures: Organizational Citizenship Behavior Checklist (OCB-C), Counterproductive Work Behavior Checklist (CWB-C) and Impression Management (IM) by Warning

	No Warning (n = 306)		Warning (n = 314)				
	M	*SD*	*M*	*SD*	*d*	df_1, df_2	*F*
OCB-C	29.84	7.85	29.48	7.38	-.05	1, 618	.34
CWB-C	16.23	5.51	15.99	5.05	-.05	1, 618	.33
IM	7.94	4.64	8.16	4.55	.05	1, 618	.34

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 32

Study 3. Fit Indices for Models Testing towards Measurement (Scalar) Equivalence for HEXACO

Model	Model Description	χ^2	df	CFI	TLI	RMSEA	SRMR
0	CFA with all data	37286.52*	2386	.75	.74	.06	.07
1	Configural equivalence	9275.49*	4772	.70	.69	.06	.08
2	Metric equivalence	9365.30*	4835	.70	.69	.06	.08
3	Scalar equivalence	9427.77*	4898	.70	.70	.06	.08

Note. * = $p < .05$

Table 33

Study 3. Criteria for Model Comparisons for Measurement Equivalence Tests

Comparison	Δdf	$\Delta\chi^2$	ΔCFI
1 vs. 2	63	89.81*	.00
2 vs. 3	63	62.47	.00

Note. * = $p < .05$

Table 34

Study 3. One-Way ANOVA of HEXACO Traits by Warning

	No Warning (n = 293)		Warning (n = 300)				
	M	*SD*	*M*	*SD*	*d*	df_1, df_2	*F*
Honesty-Humility	3.92	.61	3.75	.66	-.27	1, 591	9.88*
Emotionality	2.74	.74	2.94	.70	.28	1, 591	10.98*
Extraversion	3.89	.75	3.66	.75	-.31	1, 591	12.97*
Agreeableness	3.89	.70	3.59	.68	-.43	1, 591	27.63*
Conscientiousness	4.26	.55	4.16	.52	-.19	1, 591	5.84*
Openness	4.05	.65	3.96	.62	-.14	1, 591	2.90^{+}

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 35

Study 3. One-Way ANOVA of Motivation to Fake and Faking Behavior by Warning

	No Warning			Warning					
	M	*SD*	*n*	*M*	*SD*	*n*	*d*	df_1, df_2	*F*
Motivation to Fake	3.48	1.08	288	3.31	1.06	298	-.16	1, 584	3.63^{+}
Faking Behavior	1.83	.75	282	1.56	.57	296	-.41	1, 576	23.10*

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 36

Study 3. Coefficient Estimates and 95% Confidence Intervals for Mediation Models

Y	Specific indirect effect			Total indirect effect	Direct effect	Total effect
	$X \rightarrow M_1 \rightarrow Y$	$X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$	$X \rightarrow M_2 \rightarrow Y$			
	a_1b_1	$a_1d_{21}b_2$	a_2b_2		c^1	$a_1b_1 + a_1 d_{21}b_2$ $a_2b_2 + c^1$
Honesty	.00 (.00, .01)	.00 (.00, .01)	.02 (.01, .04)	.02 (.01, .04)	-.11* (-.16, -.06)	-.09* (-.14, -.03
Emotionality	.01 (.00, .02)	.00 (.00, .00)	-.01 (-.02, .00)	.00 (-.02, .01)	.10 *(.04, .16)	.10* (.04, .16)
Extraversion	-.02 (-.04, .00)	.00 (.00, .00)	.02 (.01, .03)	.00 (-.03, .02)	-.11* (-.17, -.05)	-.11* (-.17, -.05
Agreeableness	-.01 (-.03, .00)	.00 (.00, .00)	.01 (.00, .03)	.00 (-.02, .02)	-.15* (-.20, -.09)	-.15* (-.21, -.09
Conscientiousness	-.01 (-.03, .00)	.00 (.00, .01)	.03 (.02, .04)	.02 (.00, .04)	-.08* (-.12, -.04)	-.06* (-.10, -.02
Openness	-.01 (-.02, .00)	.00 (.00, .01)	.02 (.01, .04)	.01 (.00, .03)	-.06* (-.11, .00)	-.04 (-.09, .01)

Note. X =Warning, M_1 = Motivation to fake, M_2 = Faking, $^+ = p < .10$, $* = p < .05$.

Table 37

Study 3. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Honesty-Humility

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Honesty-Humility)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	-.11	.03	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	-.04	.02	> .10
M_2 (Faking)									b_2	-.16	.04	*
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	4.23	.10	*
		R^2 = .01				R^2 = .09				R^2 = .05		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 10.82$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 38

Study 3. Regression Results Mediation by Motivation to Fake and Faking for Warning on Emotionality

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Emotionality)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	.10	.03	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	-.09	.03	*
M_2 (Faking)									b_2	.07	.05	> .10
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	3.02	.11	*
		R^2 = .01				R^2 = .09				R^2 = .04		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 7.17$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 39

Study 3. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Extraversion

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Extraversion)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	-.11	.03	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	.24	.03	*
M_2 (Faking)									b_2	-.14	.05	*
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	3.19	.11	*
		R^2 = .01				R^2 = .09				R^2 = .14		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 30.07$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 40

Study 3. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Agreeableness

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Agreeableness)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	-.15	.03	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	.18	.03	*
M_2 (Faking)									b_2	-.11	.04	*
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	3.32	.11	*
		R^2 = .01				R^2 = .09				R^2 = .12		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 25.48$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 41

Study 3. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Conscientiousness

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Conscientiousness)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	-.08	.02	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	.14	.02	*
M_2 (Faking)									b_2	-.24	.03	*
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	4.15	.08	*
		R^2 = .01				R^2 = .09				R^2 = .14		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 31.82$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 42

Study 3. Regression Results for Mediation by Motivation to Fake and Faking for Warning on Openness

		Consequent										
		M_1 (Motivation to Fake)				M_2 (Faking)				Y (Openness)		
Antecedent		Coefficient	*SE*	*p*		Coefficient	*SE*	*P*		Coefficient	*SE*	*p*
X (Warning)	a_1	-.08	.04	+	a_2	-.12	.03	*	c^1	-.06	.03	*
M_1 (Motivation to Fake)					d_{21}	.14	.03	*	b_1	.11	.02	*
M_2 (Faking)									b_2	-.18	.04	*
Constant	i_{M1}	3.39	.04	*	i_{M2}	1.23	.09	*	i_Y	3.95	.10	*
		R^2 = .01				R^2 = .09				R^2 = .06		
		$F(1, 574) = 3.43$, $p < .10$				$F(2, 573) = 26.69$, $p < .05$				$F(3, 572) = 11.34$, $p < .05$		

Note. + = $p < .10$, * = $p < .05$.

Table 43

Study 3. Correlations between Impression Management (IM) and HEXACO across Warning and No-Warning Conditions

	No Warning (n = 293)	Warning (n = 300)
Honesty-Humility[1]	.38*	.48*
Emotionality	-.07	-.19*
Extraversion	.09	.15*
Agreeableness	.32*	.33*
Conscientiousness	.25*	.31*
Openness	.12*	.14*

Note. $^{+}$ = $p < .10$, * = $p < .05$; [1] = interaction term $p < .10$, [2] = interaction term $p < .05$.

Table 44

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Honesty-Humility

		Model 1 R^2 = .20, *MSE* = .33			Model 2 R^2 = .21, *MSE* = .33		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.35	.05	70.79*	3.35	.05	70.92*
IM (*X*)	b_1	.06	.01	11.73*	.06	.01	11.75*
Warning (*M*)	b_2	-.09	.02	-3.81*	-.17	.05	-3.57*
X × *M*	b_3				.01	.01	1.93^+

Note. $^+ = p < .10$, * = $p < .05$.

Table 45

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Emotionality

		Model 1 R^2 = .04, *MSE* = .51			Model 2 R^2 = .04, *MSE* = .51		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.00	.06	50.75*	3.00	.06	50.80*
IM (*X*)	b_1	-.02	.01	-3.06*	-.02	.01	-3.06*
Warning (*M*)	b_2	.10	.03	3.42*	.17	.06	2.91*
X × *M*	b_3				-.01	.01	-1.40

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 46

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Extraversion

		Model 1 R^2 = .04, *MSE* = .56			Model 2 R^2 = .04, *MSE* = .56		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.62	.06	58.50*	3.62	.06	58.46*
IM (*X*)	b_1	.02	.01	2.94*	.02	.01	2.94*
Warning (*M*)	b_2	-.11	.03	-3.70*	-.16	.06	-2.51*
X × *M*	b_3				.01	.01	.77

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 47

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Agreeableness

		Model 1 R^2 = .15, *MSE* = .42			Model 2 R^2 = .15, *MSE* = .43		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.35	.05	62.16*	3.35	.05	62.10*
IM (*X*)	b_1	.05	.01	8.33*	.05	.01	8.32*
Warning (*M*)	b_2	-.16	.03	-5.78*	-.16	.05	-2.96*
$X \times M$	b_3				.00	.01	.10

Note. $^+ = p < .10$, * = $p < .05$.

Table 48

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Conscientiousness

		Model 1 R^2 = .09, *MSE* = .27			Model 2 R^2 = .09, *MSE* = .27		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.95	.04	92.73*	3.95	.04	92.66*
IM (*X*)	b_1	.03	.01	7.09*	.03	.01	7.08*
Warning (*M*)	b_2	-.06	.02	-2.71*	-.08	.04	-1.84^{+}
$X \times M$	b_3				.00	.01	.57

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 49

Study 3. Regression Results for Interaction between Warning and Impression Management to Predict Openness

		Model 1 R^2 = .02, *MSE* = .40			Model 2 R^2 = .02, *MSE* = .40		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	3.86	.05	74.13*	3.86	.05	74.06*
IM (*X*)	b_1	.02	.01	3.19*	.02	.01	3.19*
Warning (*M*)	b_2	-.05	.03	-1.80$^+$	-.06	.05	-1.11
X × *M*	b_3				.00	.01	.25

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 50

Study 3. Correlations between Organizational Citizenship Behavior and HEXACO across Warning and No-Warning Conditions

	No Warning (n = 293)	Warning (n = 300)
Honesty-Humility	.07	.00
Emotionality	-.10	-.19*
Extraversion	.26*	.32*
Agreeableness	.11^{+}	.10^{+}
Conscientiousness	.17*	.24*
Openness	.18*	.19*

Note. $^{+}$ = $p < .10$, * = $p < .05$; ^{1} = interaction term $p < .10$, ^{2} = interaction term $p < .05$.

Table 51

Study 3. Regression Results for Interaction between Honesty-Humility and Warning to Predict Organizational Citizenship Behavior

		Model 1			Model 2		
		R^2 = .00, *MSE* = 57.20			R^2 = .00, *MSE* = 57.30		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	28.14	1.91	14.76*	27.96	1.92	14.58*
Honesty-Humility (*X*)	b_1	.37	.49	.76	.41	.49	.84
Warning (*M*)	b_2	-.15	.31	-.48	1.43	1.92	.75
X × *M*	b_3				-.41	.49	-.84

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 52

Study 3. Regression Results for Interaction between Emotionality and Warning to Predict Organizational Citizenship Behavior

		Model 1 R^2 = .02, *MSE* = 56.17			Model 2 R^2 = .02, *MSE* = 56.12		
		Coefficient	*SE*	*T*	*Coefficient*	*SE*	*t*
Intercept	i_1	33.82	1.26	26.93*	33.94	1.26	26.95*
Emotionality (*X*)	b_1	-1.50	.43	-3.49*	-1.52	.43	-3.54*
Warning (*M*)	b_2	-.03	.31	-.11	1.46	1.26	1.16
X × *M*	b_3				-.53	.43	-1.22

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 53

Study 3. Regression Results for Interaction between Extraversion and Warning to Predict Organizational Citizenship Behavior

		Model 1 R^2 = .08, *MSE* = 52.53			Model 2 R^2 = .09, *MSE* = 52.59		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	18.59	1.53	12.19*	18.61	1.53	12.19*
Extraversion (*X*)	b_1	2.91	.39	7.35*	2.91	.40	7.34*
Warning (*M*)	b_2	.14	.30	.48	-.60	1.53	-.39
X × *M*	b_3				.20	.40	.50

Note. $^+ = p < .10$, * = $p < .05$.

Table 54

Study 3. Regression Results for Interaction between Agreeableness and Warning to Predict Organizational Citizenship Behavior

		Model 1			Model 2		
		R^2 = .01, *MSE* = 56.74			R^2 = .01, *MSE* = 56.84		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	25.41	1.71	14.85*	25.40	1.71	14.82*
Agreeableness (*X*)	b_1	1.11	.45	2.47*	1.11	.45	2.47*
Warning (*M*)	b_2	-.02	.32	-.05	.24	1.71	.14
X × *M*	b_3				-.07	.45	-.15

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 55

Study 3. Regression Results for Interaction between Conscientiousness and Warning to Predict Organizational Citizenship Behavior

		Model 1			Model 2		
		R^2 = .04, *MSE* = 54.96			R^2 = .04, *MSE* = 54.98		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	17.49	2.42	7.24*	17.43	2.42	7.21*
Conscientiousness (*X*)	b_1	2.87	.57	5.04*	2.89	.57	5.07*
Warning (*M*)	b_2	-.03	.31	-.09	-2.24	2.42	-.93
X × *M*	b_3				.53	.57	.92

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 56

Study 3. Regression Results for Interaction between Openness and Warning to Predict Organizational Citizenship Behavior

		Model 1			Model 2		
		R^2 = .03, *MSE* = 55.46			R^2 = .03, *MSE* = 55.56		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	20.95	1.96	10.71*	20.95	1.96	10.69*
Openness (*X*)	b_1	2.15	.48	4.46*	2.15	.48	4.45*
Warning (*M*)	b_2	-.09	.31	-.28	-.23	1.96	-.12
X × *M*	b_3				.04	.48	.08

Note. $^+ = p < .10$, * = $p < .05$.

Table 57

Study 3. Correlations between Counterproductive Work Behavior and HEXACO across Warning and No-Warning Conditions

	No Warning (n = 293)	Warning (n = 300)
Honesty-Humility[2]	-.18*	-.39*
Emotionality	.05	.18*
Extraversion	-.20*	-.20*
Agreeableness	-.32*	-.41*
Conscientiousness	-.29*	-.28*
Openness	-.21*	-.16*

Note. [+] = $p < .10$, * = $p < .05$; [1] = interaction term $p < .10$, [2] = interaction term $p < .05$.

Table 58

Study 3. Regression Results for Interaction between Honesty-Humility and Warning to Predict Counterproductive Work Behavior

		Model 1			Model 2		
		R^2 = .08, *MSE* = 24.17			R^2 = .08, *MSE* = 24.02		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	24.96	1.23	20.18*	24.66	1.24	19.86*
Honesty-Humility (*X*)	b_1	-2.34	.32	-7.35*	-2.27	.32	-7.14*
Warning (*M*)	b_2	-.30	.20	-1.46	2.38	1.24	1.92^+
$X \times M$	b_3				-.70	.32	-2.18*

Note. $^+ = p < .10$, * = $p < .05$.

Table 59

Study 3. Regression Results for Interaction between Emotionality and Warning to Predict Counterproductive Work Behavior

		Model 1			Model 2		
		R^2 = .01, *MSE* = 26.04			R^2 = .02, *MSE* = 25.97		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	13.68	.86	16.00*	13.57	.86	15.84*
Emotionality (*X*)	b_1	.81	.29	2.79*	.84	.29	2.87*
Warning (*M*)	b_2	-.19	.21	-.88	-1.54	.86	-1.80^{+}
X × *M*	b_3				.48	.29	1.63

Note. $^{+}$ = $p < .10$, * = $p < .05$.

Table 60

Study 3. Regression Results for Interaction between Extraversion and Warning to Predict Counterproductive Work Behavior

		Model 1 R^2 = .04, *MSE* = 25.33			Model 2 R^2 = .04, *MSE* = 25.37		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	21.15	1.06	19.97*	21.15	1.06	19.94*
Extraversion (*X*)	b_1	-1.37	.28	-4.96*	-1.37	.28	-4.96*
Warning (*M*)	b_2	-.26	.21	-1.24	-.39	1.06	-.37
X × *M*	b_3				.03	.28	.12

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 61

Study 3. Regression Results for Interaction between Agreeableness and Warning to Predict Counterproductive Work Behavior

		Model 1			Model 2		
		R^2 = .13, *MSE* = 22.93			R^2 = .13, *MSE* = 22.91		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	26.08	1.09	23.97*	26.05	1.09	23.94*
Agreeableness (*X*)	b_1	-2.70	.29	-9.43*	-2.70	.29	-9.45*
Warning (*M*)	b_2	-.51	.20	-2.52*	.74	1.09	.68
X × *M*	b_3				-.33	.29	-1.17

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 62

Study 3. Regression Results for Interaction between Conscientiousness and Warning to Predict Counterproductive Work Behavior

		Model 1			Model 2		
		R^2 = .08, *MSE* = 24.25			R^2 = .08, *MSE* = 24.29		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	27.46	1.61	17.11*	27.45	1.61	17.09*
Conscientiousness (*X*)	b_1	-2.72	.38	-7.20*	-2.72	.38	-7.20*
Warning (*M*)	b_2	-.25	.20	-1.24	-.52	1.61	-.32
X × *M*	b_3				.06	.38	.17

Note. $^+$ = $p < .10$, * = $p < .05$.

Table 63

Study 3. Regression Results for Interaction between Openness and Warning to Predict Counterproductive Work Behavior

		Model 1 R^2 = .03, *MSE* = 25.49			Model 2 R^2 = .04, *MSE* = 25.52		
		Coefficient	*SE*	*t*	*Coefficient*	*SE*	*t*
Intercept	i_1	21.96	1.33	16.55*	21.96	1.33	16.53*
Openness (*X*)	b_1	-1.49	.33	-4.55*	-1.49	.33	-4.54*
Warning (*M*)	b_2	-.17	.21	-.83	-.88	1.33	-.66
X × *M*	b_3				.18	.33	.54

Note. $^+$ = $p < .10$, * = $p < .05$.

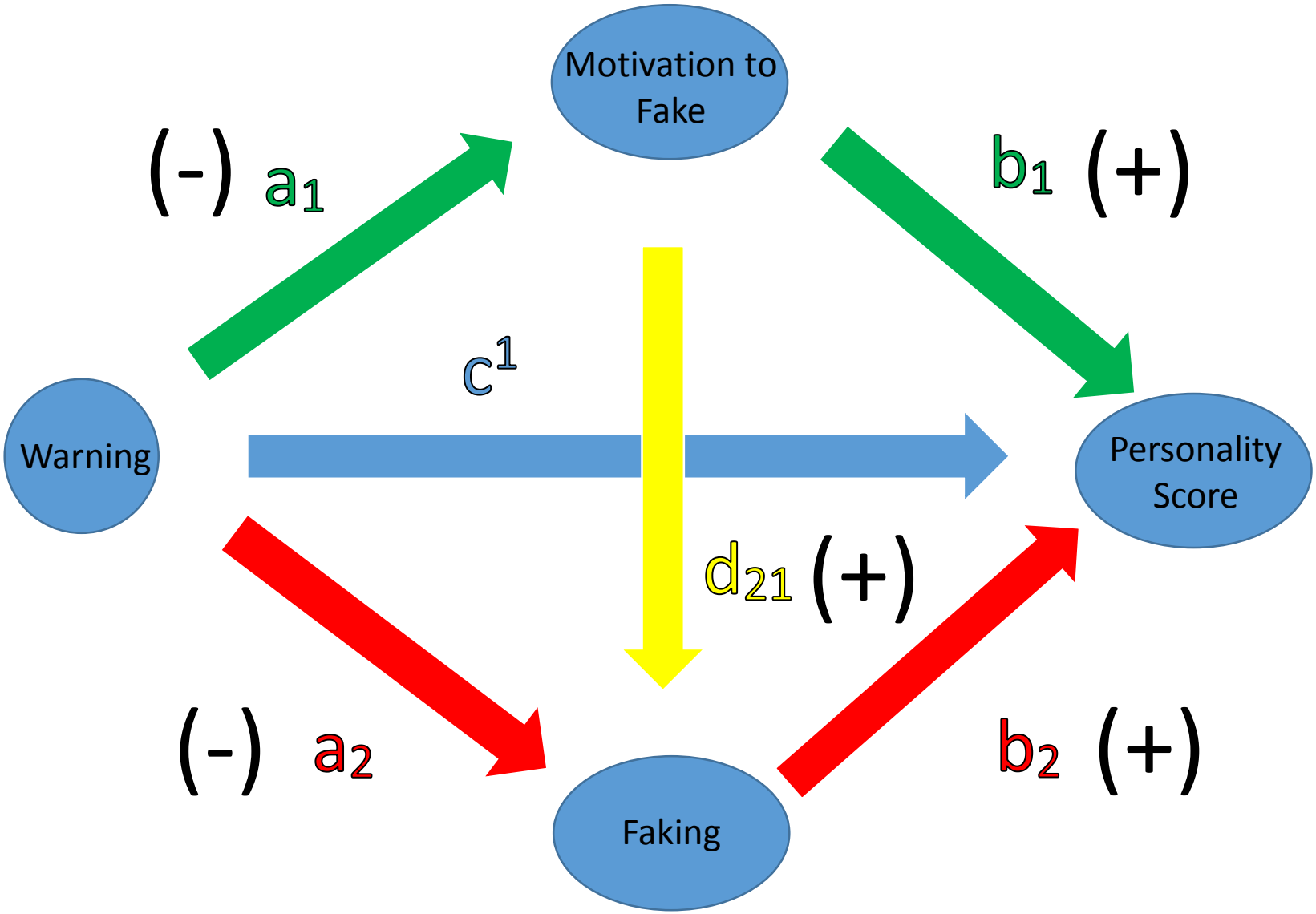

Figure 1. Mediation Model.

References

Adair, C. (2014). *Interventions for addressing faking on personality assessments for employee selection: A Meta-analysis.* Retrieved from http://via.library.depaul.edu/csh_etd/93/

AERA, APA. NCME. (1999). Standards for educational and psychological testing.

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

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

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and assessment, 9*(1-2), 9-30.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Bentler, P.M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two-cut score strategies. *Personnel Psychology, 62*(4), 833-863.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317-335.

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment, 23*(2), 532-544.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071.

Brannick, M.T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*, 201–213.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* Guilford Publications.

Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance, 24*(4), 358-372.

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.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: An Interdisciplinary Journal, 9*, 233–255.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*(3), 267-307.

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*(4), 847-860.

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16*(2), 155-169.

Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment, 17*(1), 47-60.

Cronbach, L. J. (1988). Five perspectives on validation argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Erlbaum

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192-203.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16*(1), 81-106.

Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology, 29*(3), 479-493.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1), 40-57.

DuVernet, A. Whelan, T., Seiler, S., & Hess, C. (2014). *Warning against faking: Investigating the measurement equivalence of personality scales.* Poster session presented at the Annual Convention of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1), 1-23.

Ellingson, J. E., Heggestad, E. D., & Makarius, E. E. (2012). Personality retesting for managing intentional distortion. *Journal of Personality and Social Psychology, 102*(5), 1063-1076.

Ellingson, J. E., & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of VIE theory. *Human Performance, 24*(4), 322-337.

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology, 97*(4), 866-880.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and assessment, 11*(1/2), 340-344.

Goffin, R. D., & Woods, D. M. (1995). Using personality testing for personnel selection: Faking and test- taking inductions. *International Journal of Selection and Assessment, 3*(4), 227-236.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review,36*(3), 341-355.

Griffith, R. L., & Peterson, M. H. (Eds.). (2006). *A closer examination of applicant faking behavior.* IAP.

Griffith, R. L., & Peterson, M. H. (2011). One piece at a time: the puzzle of applicant faking and a call for theory. *Human Performance, 24*(4), 291-301.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408-420.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Press.

Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*(1), 9-24.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*(5), 1270-1285.

Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality*, *29*(3), 362-372.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, *11*(2-3), 209-244.

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, *1*(3), 272-290.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: An Interdisciplinary Journal*, *6*, 1–55.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, *13*(4), 371-388.

Jansen, A., König, C. J., Stadelmann, E. H., & Kleinmann, M. (2012). Applicants' self-presentational behavior: What do recruiters expect and what do they get? *Journal of Personnel Psychology*, *11*(2), 77-85.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance:

integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875-925.

Kelloway, E.K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior*, *16*, 215–224.

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology*, *93*(1), 140-154.

König, C. J., Hafsteinsson, L. G., Jansen, A., & Stadelmann, E. H. (2011). Applicants' Self-presentational Behavior across Cultures: Less self-presentation in Switzerland and Iceland than in the United States. *International Journal of Selection and Assessment*, *19*(4), 331-339.

König, C. J., Wong, J., & Cen, G. (2012). How much do Chinese applicants fake? *International Journal of Selection and Assessment*, *20*(2), 247-250.

Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses.*International Journal of Selection and Assessment*, *15*(2), 220-231.

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*(1), 202-210.

Marcus, B. (2009). ‘Faking’ from the applicant's perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment, 17*(4), 417-430.

McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment, 11*(4), 265-276.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*(5), 812-821.

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*(4), 979-1016.

McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment, 78*(2), 348-369.

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361-388.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance.* Routledge.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60(*4), 1029-1049.

ONet Online (n.d.). *Summary Report for 41-2031.00 – Retail Salespersons.* Retrieved from http://www.onetonline.org/link/summary/41-2031.00

Pace, V. L., & Borman, W. C. (2006). The use of warning to discourage faking on noncognitive inventories. In Syvantek, D., & Griffiths, R. L. (Eds.) *A closer examination of applicant faking behavior* (pp.283-304). IAP.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.

Peeters, M. A., Tuijl, H. F., Rutte, C. G., & Reymen, I. M. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality, 20*(5), 377-396.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*(1), 153-172.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-529.

Robson, S. M., Jones, A., & Abraham, J. (2007). Personality, faking, and convergent validity: a warning concerning warning statements. *Human Performance*, *21*(1), 89-106.

Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359-371.

Sackett, P. R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance, 24*(4), 379-385.

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 91*(3), 613-621.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, *78*(6), 966-974.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*, 117-125.

Spector, P. E., Bauer, J. A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: do we know what we think we know? *Journal of Applied Psychology, 95*(4), 781-790.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, *9*(2), 219-242.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180.

Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A “multisaturation” perspective on faking as performance. *Human Performance*,*24*(4), 302-321.

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), 301-316.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*(3), 175-199.

Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability?. *Journal of Applied Psychology*, *90*(2), 306-322.

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

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods*, *7*(2), 168-190.

Ziegler, M., McCann, C., & Roberts, R. D. (2012). *New perspectives on faking in personality assessments*. OUP.

Ziegler, M., Maaß, U., Griffith, R., & Gammon, A. (in press). What is the nature of faking? Modeling distinct response patterns and quantitative differences in faking at the same time. *Organizational Research Methods*.