The Role of Emotions as Mechanisms of Mid-Test Warning Messages during Personality Testing: A Field Experiment

Hairong Li
Renmin University of China

Jinyan Fan
Auburn University

Guoxiang Zhao & Minghui Wang
Henan University

Lu Zheng
Huazhong University of Science & Technology

Hui Meng
East China Normal University

Qingxiong (Derek) Weng
University of Science and Technology of China

Yanping Liu
Sun Yet-Sen University

Filip Lievens
Singapore Management University

Accepted at the Journal of Applied Psychology.

© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/apl0000885

Author Note

Hairong Li, School of Labor and Human Resources, Renmin University of China; Jinyan Fan, Department of Psychology, Auburn University; Guoxiang Zhao, Department of Psychology, Henan University; Minghui Wang, Department of Psychology, Henan University; Lu Zheng, School of Management, Huazhong University of Science & Technology; Hui Meng, School of Cognitive and Psychological Sciences, East China Normal University; Qingxiong (Derek) Weng, Department of Business Administration, University of Science and Technology of China; Yanping Liu, Department of Psychology, Sun Yet-Sen University; Filip Lievens, Lee Kong Chian School of Business, Singapore Management University.
Hairong Li and Jinyan Fan contributed equally to this manuscript. The study was conducted and the initial draft was prepared when Hairong Li was at the Department of Psychology of Auburn University, and revisions were completed after Hairong Li joined the Renmin University of China. This research was partially supported by a small grant from the Society for Industrial and Organizational Psychology awarded to Jinyan Fan, and a Key Research Project of Philosophy and Social Science Research, Ministry of Education, China (No.13JZDW010), awarded to Guoxiang Zhao.

We thank Wu Liu for referring us to the broad emotion literature. We also thank the Graduate School of Henan University for data collection support.

Correspondence concerning this article should be addressed to Guoxiang Zhao, Department of Psychology, Henan University, 85 Minlun Street, Kaifeng, Henan, P. R. China, 475001. Electronic mail may be sent to zhaogx@henu.edu.cn.
Abstract

This study focuses on the role of emotions in personnel selection and faking research. In particular, we posit that emotions are likely to be activated when applicants receive warning messages from organizations. Drawing on Nabi’s (1999) cognitive-functional model of discrete negative emotions, we propose and empirically test the effects of three discrete negative emotions (guilt, fear, and anger) triggered by a warning message during a personality test on personality score accuracy and perceived test fairness. Participants in this within-subjects field experiment were 1,447 applicants for graduate school at a large public university in China. They completed two parallel forms of a personality test: one within a selection context, and another within a developmental context 6 months later as a baseline measure. In the selection context, a warning (or a control) message was randomly assigned to participants during the personality test. Emotions and perceived test fairness were measured after the test was completed. Results indicated that guilt, fear, and anger each played a unique role. Guilt explained how mid-test warnings improved personality score accuracy among fakers, whereas fear accounted for why non-fakers over-corrected their personality scores. Finally, anger explained why the mid-test warnings reduced perceived test fairness for both fakers and non-fakers. Theoretical and practical implications are discussed.

Keywords: warnings, guilt, fear, anger, applicant faking, applicant perceptions
The Role of Emotions as Mechanisms of Mid-Test Warning Messages during Personality Testing: A Field Experiment

Although personality measures have become a popular tool for personnel selection during the last three decades (e.g., Barrick & Mount, 1991; Hough & Oswald, 2008; Judge, Rodell, Klinger, Simon, & Crawford, 2013; Hough, Oswald, & Ock, 2015), many organizations are concerned that self-reported personality measures are prone to faking or response distortion (e.g., Hough & Oswald, 2008; Rosse, Stecher, Levin, & Miller, 1998; Sackett, Lievens, Van Iddekinge, & Kuncel, 2017). Faking is defined as “the tendency to deliberately present oneself in a more positive manner than is accurate, in order to meet the perceived demands of the testing situation” (Fan et al., 2012, p. 867). Despite ongoing debates among selection scholars on whether faking is a problem in real-world selection contexts (see Griffith & Peterson, 2006; Ziegler, MacCann, & Roberts, 2012, for comprehensive treatises), applicant faking remains an important issue that needs further research, if applied psychologists are to advocate wider use of personality tests in employment settings.

Over the years, numerous faking models have been proposed (e.g., Goffin & Boyd, 2009; Shoss & Strube, 2011), which have substantially enhanced our understanding of the antecedents and processes of applicant faking. However, these faking models are mostly cognition-based, overlooking the role of emotions. Given the well-established notion that emotions can shape individuals’ motivational states and behaviors (e.g., Lerner, Li, Valdesolo, & Kassam, 2015; Parker, Bindl, & Strauss, 2010) and the robust empirical evidence showing the relevance of emotions in various organizational behavior contexts (e.g., Duffy & Shaw, 2000; Fox & Spector, 2000; Lerner et al., 2015), it is surprising that theoretical work, as well as empirical studies, on the role of emotions in applicant faking are scarce.
We argue that emotions are relevant to research on selection and faking, and that emotions might be activated when applicants receive warnings from organizations. Scholars have long recognized that the social relationship between the applicant and the hiring organization embedded within the selection setting is regulated by a set of moral rules (Herriot, 1989; Ramsay, Gallois, & Callan, 1997). Applicant faking and warnings may be interpreted as justifiable moral transgressions or moral actions, depending on the perspective taken (the applicant vs. the hiring organization) and the faker status (fakers vs. non-fakers). Within this context, some emotions may act as evaluators of and regulators for moral behaviors, and thus may motivate applicants to conform to or challenge the moral rules.

To our knowledge, Ellingson, Heggestad, and Markarius (2012) conducted the only study in which emotions were examined in the context of applicant faking and warnings. These authors conducted a lab study and discovered that state guilt was positively related to an increase in personality test score accuracy among fakers receiving the retest warnings, but not among fakers receiving a control message. Guilt arose because warned fakers were made aware that their behaviors had violated the moral rules between the applicant and the test administrator, and elevated guilt then motivated warned fakers to make reparation for their fault (Lazarus, 1991). Ellingson et al.’s (2012) seminal work underlined the potential role of emotions in selection and faking, thereby at the same time calling for much needed additional research focusing on multiple emotions, on fakers and non-fakers, and on test scores as well as applicant perceptions. For instance, Ellingson et al. (2012) also reported that non-fakers tended to underreport their personality scores after receiving the retest warnings, but did not examine the underlying mechanisms. We suggest that fear of potential negative consequences might explain this finding. Fear can be aroused by threats to one’s physical or psychological self (i.e., receiving...
Emotions, Warnings, and Faking

a warning) and prompts non-fakers to escape from the threatening agent by underreporting their personality scores (Lazarus, 1991). In addition, feelings of anger might also be a relevant emotion, which might be solicited when applicants feel they are unjustifiably warned. Driven by the tendency to correct the wrongdoing (e.g., Ford, Wang, Jin, & Eisenberger, 2018), anger might prompt non-compliance and even opposing responses (e.g., Dillard, Plotnick, Godbold, Freimuth, & Edgar, 1996; Milberg & Clark, 1988). Feelings of anger toward the warning may also bring out negative evaluations of the testing system, resulting in lowered perceived test fairness. In summary, including multiple negative emotions in theorizing about faking and warning processes seem fruitful for advancing our understanding of applicant faking in selection.

The purpose of the present research is to investigate several discrete negative emotions as mechanisms of mid-test warnings in affecting applicant faking and perceptions on an online personality test within a selection context. Mid-test warnings entail sending test-takers a warning message in the middle of the test (to be discussed shortly). We focus on mid-test warnings for two reasons. First, mid-test warnings have emerged as a new, promising faking-mitigation procedure (e.g., Ellingson et al., 2012; Fan et al., 2012), whose mechanisms, however, are still poorly understood. Second, mid-test warnings represent a potentially significant event/stimulus to test-takers, and are expected to generate several specific emotions probably in moderate strength, which might explain the warning effect on applicant faking and perceptions. In other words, mid-test warnings provide an excellent context to examine the role of emotions on applicant faking and perceptions. Figure 1 depicts our conceptual model. As can be seen, mid-test warnings are posited to evoke three negative emotions of guilt, fear, and anger, which in turn influence applicant faking (i.e., personality score accuracy) and applicant perceptions (i.e.,
perceived test fairness). We also posit that the strength of some of these pathways might differ between fakers and non-fakers.

The present research aims to make several contributions to selection research. First, in the most general sense, our study is among the first to examine emotions in the process of employee testing for selection, and shows that emotions may have a nontrivial influence on applicant faking. Second, this study advances our understanding of how mid-test warnings, as a specific faking-mitigation strategy, work via several emotional mechanisms. We show that different emotions play unique roles among fakers and non-fakers. This extends previous research of Ellingson et al. (2012), which focused only on guilt as an emotional mechanism and only on fakers. Third, the present study reveals the emotional pathways through which applicant perceptions are formed. Previous research mostly concentrated on the nature and consequences of applicant perceptions, paying little attention to their formative process (for an exception, see the Applicant Attribution-Reaction Theory of Ployhart and Harold, 2004, but their model is cognition-based rather than affect-based).

**An Introduction to Mid-test Warnings**

To address applicant faking on personality tests, numerous faking-mitigation strategies have been developed (see Burns & Christiansen, 2011; Kuncel & Borneman, 2007, for reviews). Of all these strategies, there has emerged a new, “test-warning-retest” approach, which typically entails the following steps: (a) administering the initial test (either the entire test or a part of it), (b) identifying fakers based on the initial test results, (c) warning fakers, and (d) offering retesting. Over the years, researchers have suggested several variants of mid-test warnings (e.g., Burns, Fillipowski, Morris, & Shoda, 2015; Butcher, Morfitt, Rouse, & Holden, 1997; Ellingson et al., 2012; Fan et al., 2012; Landers, Sackett, & Tuzinski, 2011). The mid-test warning
procedure developed by Fan et al. (2012) blends the strengths of similar procedures. In Fan et al.’s procedure (see Figure 2), a test-taker first completes the “initial block,” which includes two faking measures: An impression management (IM) scale (Paulhus, 1991) and a bogus statement (BS) scale (Dwight & Donavan, 2003). If a test-taker’s score in either the IM or BS scale exceeds the pre-set criterion, he/she is flagged as a faker, and receives a warning message on the screen; otherwise, he/she is categorized as a non-faker, and receives a control message. Then all test-takers proceed to the “main block,” which includes the same BS and IM items, and the whole set of personality items. As the warning is delivered in the middle of the test or after the initial test rather than at the beginning of the test, we refer to this new approach as “mid-test warnings,” to differentiate them from the more traditional pre-test warnings.

A few empirical studies documented several effects associated with mid-test warnings. First, mid-test warnings are effective in lowering personality scores (e.g., Butcher et al., 1997; Fan et al., 2012; Landers et al., 2011). Second, mid-test warnings seem to increase personality score accuracy for fakers, but may also decrease score accuracy for non-fakers (e.g., Ellingson et al., 2012), who tend to over-correct their scores after being warned. However, as noted earlier, the reasons for these findings are poorly understood. Third, mid-test warnings may lead to negative applicant reactions (e.g., Fan et al., 2012). Again, we do not know exactly why this is the case. Therefore, it seems fair to conclude that despite initial promising evidence for mid-test warnings as an effective faking-mitigation procedure, we know little about the potential underlying mechanisms of how mid-test warnings resort effects on applicant faking (see Ellingson et al. 2012, for the sole exception) and applicant perceptions. As noted above, we posit that emotions represent unexplored mechanisms underlying mid-test warnings.

**Emotions as Mechanisms of Mid-test Warnings**
Current Research Design

To facilitate hypothesis development, a brief description of the current research design is in order. In this within-subjects field experiment, applicants completed two parallel forms of a personality test, the first time within a selection context (Form A), and the second time within a developmental context six months later (Form B). In the selection context, a mid-test warning or a control message was randomly assigned to applicants after the initial block. The developmental context personality data was obtained as a baseline measure. We define personality score inaccuracy as the distance between scores reported in the selection context and those reported in the developmental context. The smaller the distance, the higher the personality score accuracy.

Emotions as Responses to Mid-test Warnings

A mid-test warning message can be viewed as a persuasive message aimed to encourage test-takers to respond honestly on subsequent personality items. Recognizing the important role emotions play in persuasion, Nabi (1999) proposed a cognitive-functional model to depict how certain discrete, message-induced negative emotions impact information processing and attitude change. Emotions are “internal, mental states presenting evaluative, valenced reactions to events, agents, or objects that vary in intensity” (Ortony, Clore, & Collins, 1988). Emotions are elicited by a target or cause, often include physiological reactions and action tendencies, and are more transient and intense than moods (Frijda, 1986; Lazarus, 1991). According to Nabi (1999), a persuasive message that incorporates an emotion’s core relational theme is likely to trigger that emotion, which in turn will motivate information processing, and ultimately lead to message acceptance or rejection. A core relational theme is the central meaning associated with a certain emotion, and it provides a summary for the relational harm or benefit underlying each specific emotion (Lazarus, 1991). Nabi (1999) suggested that five negative emotions can be potentially
evoked by persuasion messages: Guilt, fear, anger, disgust, and sadness. We argue below that the emotions of guilt, fear, and anger are most likely to be induced by the mid-test warnings used in the current study, because it embodies the core relational themes of these emotions.\(^1\)

**Guilt.** Guilt is a social emotion arising from one’s violation of a moral, ethical, or cultural standard and rule (Ausubel, 1955; Izard, 1977; Lazarus, 1991; Tangney, 1999). According to Lazarus (1991), the core relational theme of guilt is “having transgressed a moral imperative” (p. 122). Selection testing situations typically involve an interpersonal relationship between the test-taker and the test administrator (and by extension, the hiring organization), and the relationship is characterized by a set of moral rules (Argyle, Furnham, & Graham, 1981; Ellingson et al., 2012). For example, the assessment results will only be used to evaluate the test-takers’ competency; in the meantime, the test-taker is supposed to respond honestly on the assessment. The current mid-test warnings contains a detection/accusation component conveying to the test-taker that “You’ve been caught faking,” which suggests a moral offense (Herriot, 1989), thus clearly embodying the core relational theme of guilt. As such, the mid-test warnings should elicit at least a modest level of guilt for fakers who have violated the aforementioned moral rules; however, such an effect should be weaker for non-fakers, because non-fakers have not violated these moral rules. Thus, we expected the mid-test warnings to trigger a stronger feeling of guilt for fakers than for non-fakers.

**Hypothesis 1:** Applicants who receive the mid-test warnings will report a higher level of guilt than those who receive the control message. Faker status will moderate this relationship such that it will be stronger for fakers than for non-fakers.

---

\(^1\) We contend that the mid-test warnings used in the current study are unlikely to evoke sadness whose core relational theme is “irrevocable loss” (Lazarus, 1991), in that our mid-test warnings represent a second chance, not an irrevocable verdict. Regarding disgust, its core relational theme is “taking in or being too close to an indigestible object or idea” (Lazarus, 1991). The mid-test warnings were developed by Fan et al. (2012) based on interpersonal justice principles, and thus should be seen as friendly instead of indigestible. We conducted a pilot study to show that our mid-test warnings neither solicited sadness nor disgust (for more details, see online supplemental materials).
Fear. Fear is elicited when a situation is perceived as both threatening to one’s physical or psychological self and out of one’s control (Frijda, 1986; Lazarus, 1991; Scherer, 1984). The core relational theme of fear is “immediate, concrete and overwhelming physical danger” (Lazarus, 1991, p. 122). In selection contexts, a job applicant’s main goal is to survive the application processes and get the job offer. As the typical mid-test warnings contains a consequence component conveying to the test-taker that continued faking, if caught, would result in negative consequences, it is conceivable that reminding one of potential negative consequences is likely to make the danger of imminent harm salient, which is the core relational theme of fear. We expected mid-test warnings to trigger at least a modest level of fear among fakers and non-fakers alike, because non-fakers, although they have not faked, may still experience the threat conveyed by the mid-test warnings.

Hypothesis 2: Applicants who receive the mid-test warnings will report a higher level of fear than those who receive the control message.

Anger. Anger is an emotion that involves an appraisal of responsibility for wrongdoing by another person or entity (Gibson & Callister, 2010). Anger can be triggered by situations where obstacles are perceived to impede goal-oriented behavior (Averill, 1982; Hampton, 1978; Izard, 1977; Plutchik, 1980), or by perceived injustice (Carver & Harmon-Jones, 2009; Gibson & Callister, 2010), or both. In selection settings, mid-test warnings may induce anger for both fakers and non-fakers because it incorporates the core relational theme of anger, “demeaning offense against me or mine” (Lazarus, 1991, p. 122), but for different reasons and to different degrees. Fakers use faking as a strategy to achieve their goal (i.e., passing the test and getting selected), and mid-test warnings can be perceived as an impediment to achieving that goal, and thus should arouse at least a modest level of anger (Nabi, 1999). For non-fakers, the unwarranted accusation that they have faked not only entails the above impediment-to-goal effect, but also
represents a moral violation on the part of the test administrator (Graham et al., 2011; Rai & Fiske, 2011), and therefore, should yield stronger feelings of anger.

**Hypothesis 3:** Applicants who receive the mid-test warnings will report a higher level of anger than those who receive the control message. Faker status will moderate this relationship such that it will be stronger for non-fakers than for fakers.

**Effects of Emotions on Changes in Personality Score Accuracy Due to Warnings**

The emotions people feel may shape the ways they respond to persuasion attempts (e.g., Frijda, 1986; Izard, 1977; Lazarus, 1991; Plutchik, 1980). Nabi’s (1999) model demonstrates that message-triggered emotions influence cognitive processing and subsequently lead to message acceptance or rejection. We elaborate below on how the three negative emotions are associated with changes in personality score accuracy from the initial (Time 1) to the main block (Time 2).

**Guilt.** A feeling of guilt leaves individuals feeling tense and regretful of their past actions (Tangney, 1999). This discomfort caused by internal blame motivates individuals to remedy the situation so that guilt may be dissipated (Izard, 1977; Lazarus, 1991; Roseman, Wiest, & Swanz, 1994). In the current context, when offered such a remedy opportunity through retesting, test-takers are likely to accept the warning privately by responding more honestly to subsequent personality items (Cialdini & Goldstein, 2004). This should lead to an increase in score accuracy from Time 1 (initial block) to Time 2 (main block). Thus, there should be a positive relationship between guilt and an increase in personality score accuracy.

Furthermore, for the fakers who are warned, the level of guilt should closely correspond to the level of motivation for remedying the situation, and thus should be positively associated with an increase in score accuracy (see Ellingson et al., 2012). However, for other test-takers (i.e., the fakers who are unwarned, the non-fakers who are warned, and the non-fakers who are unwarned), the level of guilt after the middle-message (warning or control) should have weaker
connotations for the motivation to remedy the situation, and thus should be less likely to be associated with an increase in score accuracy (Ellingson et al., 2012). Therefore, we expected faker status to moderate the link between guilt and an increase in personality score accuracy, with the link being stronger for fakers than for non-fakers. Together, we hypothesize that guilt should mediate the effect of mid-test warnings on an increase in personality score accuracy, with faker status moderating both the treatment (warning vs. control messages) – guilt link (H1) and the guilt – an increase in score accuracy link.

**Hypothesis 4:** Faker status will moderate the relationship between guilt and an increase in personality score accuracy such that the relationship will be stronger for fakers than for non-fakers (H4a). Faker status will also moderate the indirect link of treatment – guilt – an increase in score accuracy such that it will be stronger for fakers than for non-fakers (H4b).

**Fear.** From an evolutionary perspective, fear is aroused as a signal of danger or threat to trigger appropriate adaptive responses for survival (Ohman & Wiens, 2003). Researchers have used the drive model to explain the effects of fear arousal and compliance (Hovland, Janis, & Kelley, 1953; Janis & Feshbach, 1953). According to the drive model, when fear is aroused the recipient will become motivated to reduce such unpleasant feelings and seek protection (Lazarus, 1991). In the current context, when the mid-test warnings recommends a solution to avoid the potential punishment, the warned test-takers are likely to comply with it. In other words, to avoid potential punishment and alleviate fear, the warned test-takers will be driven to report less favorable personality scores in the retest (the main block). In this case, test-takers may accept the warning publicly, but not necessarily privately. In fact, the strategy of scaring people into compliance with an advocated course of action has been used to promote adults’ healthy behaviors (Breckler, 1993) and guide children’s appropriate behaviors (Gershoff, 2002).

As noted earlier, both fakers and non-fakers are likely to experience similar levels of fear, which in turn should lead to lowered personality scores. However, such an effect should result in
an increase in score accuracy among fakers, because they have faked and will report more accurate scores after the warning. In contrast, such an effect might result in a decrease in score accuracy among non-fakers, because they did not fake, but are forced to lower their scores. Thus, we expected faker status to moderate the relationship between fear and an increase in personality score accuracy. Together, we propose the following moderated mediation hypothesis.

**Hypothesis 5:** Faker status will moderate the relationship between fear and an increase in personality score accuracy such that the relationship will be positive for fakers, but negative for non-fakers (H5a). Faker status will moderate the indirect link of treatment – fear – an increase in personality score accuracy such that it will be positive for fakers, but negative for non-fakers (H5b).

**Anger.** Anger is often followed by a tendency to take action to correct or punish the transgressor in order to restore justice (e.g., Carver & Harmon-Jones, 2009; Ford et al., 2018; Haidt, 2003; Novaco, 1986). Empirical studies on compliance consistently found a negative association between state anger and compliance (e.g., Dillard et al., 1996; Milberg & Clark, 1988). Studies also demonstrated that anger not only decreases compliance, but also leads to changes in the individual’s behavior in a direction opposite to that advocated by the experimenter’s message (e.g., Milberg & Clark, 1988). We thus expected anger to motivate warned test-takers not to comply with the warning message (rejecting the warning), but with different behavioral patterns for fakers and non-fakers. Specifically, anger will likely prompt fakers to challenge the accusation by continuing their previous response pattern in the retest (main block), resulting in little change in personality score accuracy. In contrast, non-fakers will likely act against the warnings by providing less accurate personality scores. Recall that Ellingson et al. (2012) found that mid-test warnings reduced score accuracy for non-fakers. We argue that fear (see above) and/or anger might explain such an effect. Together, we propose the following hypothesis.
Hypothesis 6: Faker status will moderate the negative relationship between anger and an increase in personality score accuracy such that the relationship will be stronger (more negative) for non-fakers than for fakers (H6a). Faker status will also moderate the negative indirect link of treatment – anger – an increase in personality score accuracy such that it will be stronger for non-fakers than for fakers (H6b).

Effects of Emotions on Perceptions of Test Fairness

An optimal faking-mitigation approach should not only demonstrate its efficacy in reducing faking, but also have minimal negative impact on applicant perceptions (aka “applicant experience”). Applicant perceptions reflect how job applicants view the selection experience and have been linked to various outcomes such as job acceptance intentions, recommendation intentions, and perceptions of organizational attractiveness (e.g., Bauer, Maertz, Dolen, & Campion, 1998; Gilliland, 1994; McCarthy, Van Iddekinge, Lievens, Kung, Sinar, & Campion, 2013). In the current study, we focus on applicants’ perceived test fairness, defined as evaluations of whether the testing system is fair or not (Gilliland, 1993).

Research shows that when making evaluative judgements, individuals have the tendency to attend to their emotional feelings, as if asking “How do I feel about it?” (Schwarz & Clore, 1983). Affect can provide information about individuals’ unconscious appraisals and allow them to learn about their implicit judgments and decisions (Schwarz & Colore, 1983). Positive or negative emotions will likely elicit mood-congruent evaluations. That is, individuals’ judgments tend to be positive when they experience positive emotions about the object, but negative when they feel negative emotions toward the object (Clore, Gasper, & Garvin, 2001; Clore & Storbeck, 2006). Thus, it seems that negative emotions (i.e., guilt, fear, and anger) triggered by the mid-test warnings will likely prompt warned test-takers to form a negative evaluation of the testing system, resulting in lowered perceived test fairness.
On the other hand, test-takers may engage in cognitive processes and behaviors that could weaken the negative association between negative emotions and perceived test fairness. For instance, we argued earlier that feelings of guilt should motivate individuals to engage in relationship reparative behaviors (Cialdini & Goldstein, 2004; Lazarus, 1991). In the current context, guilt induced by the mid-test warnings may prompt applicants to provide positive evaluations of the testing system as a way to restore the relationship with the test administrator, in addition to lowering their personality scores. Fear may also elevate participants’ evaluations of test fairness. As argued earlier, when fear is aroused one will become motivated to reduce the unpleasant feelings and seek protection (Janis & Feshbach, 1953; Lazarus, 1991). In the current context, warned applicants might choose to report higher levels of perceive test fairness as a way to appease the test administrator, in order to avoid potential punishment. With regard to anger, studies consistently documented a negative relationship between anger and fairness perception (e.g., Pillutla & Murnighan, 1996; Xiao & Houser, 2005). However, we wanted to test whether the previous finding holds in the mid-test warnings context.

Further, faker status might act as a moderator in this situation according to previous findings that mid-test warnings had differential effects on applicant perceptions for applicants engaging in different levels of faking (e.g., Burns et al., 2015). As discussed earlier, mid-test warnings may trigger different emotions among fakers vs. non-fakers, and these emotions, in turn, might exert different effects on perceived test fairness among fakers vs. non-fakers. For example, non-fakers might feel more anger after receiving the mid-test warnings than fakers, and then provide lower test fairness evaluations. Thus, given the unclear nature of the relationships between the three negative emotions and perceived test fairness, we refrain from proposing research hypotheses; instead, we put forward the following research questions.
**Research question 1:** How will negative emotions (guilt, fear, and anger) be related to perceived test fairness?

**Research question 2:** Will negative emotions (guilt, fear, and anger) mediate the relationship between treatment (mid-test warning vs. control message) and perceived test fairness?

**Research question 3:** Will faker status moderate the indirect links of treatment – negative emotions – perceived test fairness?

### Method

**Sample and Procedure**

Participants were applicants seeking admission into graduate programs at Henan University (河南大学), a large public university located in a central city in China. Applicants who had successfully passed the nationwide graduate school entrance examinations (n = 2,482) were invited for a campus interview. These applicants were instructed to complete an online psychological assessment (which included a cognitive ability test\(^2\) and a personality test) and were given one week to complete the assessment before reporting to the campus interview. During the campus interview, which was the last hurdle in the selection process, applicants took an English language test and several subject exams, and then went through an interview. Whereas the nationwide graduate school entrance examinations have been highly competitive (eliminating 70% – 75% of applicants), the campus interviews typically have a modest elimination rate of around 30%.

**Field experiment.** Figure 2 presents the research design flow chart with a focus on the personality test. According to Figure 2, during the online personality test, either a control or

---

\(^2\) The cognitive ability test was the 30-item, timed (20 minutes) abstract reasoning subtest taken from the Chinese version of the Applied Reasoning Test (Page, 2015). Preliminary analysis indicated that cognitive ability scores did not moderate the treatment effects on any of the mediators (three negative emotions) and any of the dependent variables (change in personality score accuracy and perceived test fairness). The detailed results are included in the online supplemental materials. As such, we did not include cognitive ability scores in subsequent analyses.
warning message (see Appendix A for message scripts) was randomly assigned after participants completed the initial block (Time 1), which included a bogus statement (BS) scale, an impression management (IM) scale, and a Conscientiousness scale from Form A of a Big Five personality inventory. Then all participants proceeded to the main block (Time 2), which contained the same BS, IM scales, and the whole set of Form A of the Big Five personality inventory including the same Conscientiousness items placed in the initial block. After finishing the personality test, participants were asked to complete several questionnaires for research purposes only. Participants were first presented the messages (warning or control) they had received earlier once again and were asked to recall their emotions after receiving the message through an emotions survey. They then completed a perceived test fairness survey. We pondered the alternative strategy of measuring emotions immediately after the middle-messages. However, we were concerned that such a strategy would not only interrupt the normal testing process, but might also trigger participant reactivity, leading to heightened attention to emotional experience, intensified affective states, and consequently, inflated emotional effects. In other words, the current measurement strategy represents a more conservative test of the hypothesized emotional effects (see also Discussion). In total, 2,033 participants provided completed data on the personality test and the research surveys, for a response rate of 81.9%.

There were two deception components in the current study. First, the psychological assessment results were actually not used to make admission decisions. Second, the warning and control messages were randomly assigned rather than being assigned to fakers and non-fakers, respectively. We took several steps to minimize potential psychological risks. First, all

---

3 It is important to note that the random assignment was done for research purposes and should not be used in real-world selection settings. In field settings the test system must assign the warning message only to fakers identified in the initial block and the control message to non-fakers.
participants were told through the last screen of the online test system that they had *successfully* passed the online assessment. This should reduce participants’ tension and worries, particularly for those who received the warnings. Second, upon reporting to the campus interview, participants received a written debrief statement that explained the study purposes and the necessity of the deceptions. Participants were given the opportunity to withdraw their data from subsequent analyses, and 52 participants chose so. A $\chi^2$ test indicated that the withdrawal status had no significant association with the treatment conditions (warnings vs. control): $\chi^2 (df = 1) = .31, p = .58$. All applicants were paid 50 CNY (approximately 7 USD) cash compensation.

Third, the research team did not share participants’ assessment data with the Graduate School.\(^4\)

**Baseline personality measure.** Six months after the online assessment (or two months after participants entered into graduate school; Time 3), we obtained the baseline personality scores of those participants who were admitted and were then enrolled at this university ($n = 1,803$, with an acceptance rate of 72.6%). At many Chinese universities including this one, new graduate students go through a mandatory psychological assessment during the first semester as part of a student’s career planning and development initiative. We worked with this university to incorporate the parallel form (Form B) of the personality test into the new graduate student assessment package. In exchange, each student received a feedback report, which contained his/her personality profile, strengths and weaknesses, the degree of fit with various job families, recommended developmental activities, etc. Among the aforementioned 1,803 new graduate students, 1,460 provided complete data on the Form B of the personality test.\(^5\) The final sample

---

\(^{4}\) The present study was approved by the Institutional Review Board of Auburn University (Protocol#: 15-134 MR 1503, titled “A Comprehensive Examination of a New Faking Mitigation Procedure for Personality Tests within Selection Contexts: A Field Experiment,” PI: Jinyan Fan) and by the graduate school of Henan University.

\(^{5}\) According to the Graduate School of Henan University, there were 2,197 new graduate students in that cohort. Among them 394 were admitted with early decisions without the nationwide entrance examination, who did not go through the campus interview, and thus did not complete the online assessment, resulting in 1,803 admitted through regular admissions. Among these 1,803 new graduate students, 173 chose to delay their graduate school, 35 decided
included 1,447 participants for whom data from both test occasions were available. In the final sample, 31% were men, with a mean age of 23.44 (SD = 1.78). All participants were Chinese.

Controlling for potential confounder effects. Given the parallel forms of the personality test, and the 6-month separation between the selection (T1/T2) and development (T3) assessment, several effects might contribute to systematic score changes. According to Ellingson, Sackett, and Connelly (2007), these effects include: (a) the faking effect (different levels of motivation between T1/T2 and T3), (b) different sets of items (Form A vs. Form B), (c) second exposure to the same test with similar items, (d) maturation and life experience, and (e) other effects. Our goal was to isolate the faking effect from all other potential confounders. This is important, because score accuracy is the key outcome variable in this study. To achieve this goal, we conducted a pilot study with a group of first-year graduate students from another public university in China (East China Normal University). These participants (n = 207) completed Form A of the personality test after they entered into the university as new graduate students. Six months later, these participants were invited to complete Form B of the personality test (n = 181; response rate = 87.4%). In both occasions, these participants were told that the assessments were for research purposes only and were instructed to respond honestly. These participants were unlikely to have engaged in faking due to the non-selection context. Thus, any mean score differences between the two time points should reflect non-faking effects combined such as parallel forms, second exposure, and maturation (Ellingson et al, 2012).

Following Ellingson et al. (2012), in the main study sample we removed the non-faking effects combined based on pilot study findings from baseline scores to isolate the faking effect.

to withdraw at the campus interview, 111 somehow did not participate in the new graduate student assessment, 11 had too much missing data at the baseline assessment, and 13 did not provide their names on the baseline assessment, thus resulting in 1,460 providing complete data on the baseline personality test.
Specifically, we first calculated mean score changes of all Big Five traits in the pilot study by subtracting Form B mean scores from Form A mean scores for respective traits, which were defined as adjustment factors. Next, we obtained adjusted baseline (T3) personality scores for all Big Five traits in the main study by subtracting the adjustment factor from respective T3 Big Five personality scores. We then operationalized score accuracy as follows:

Score Accuracy at Time 1 = 5 – | Time 1 Score – Adjusted Baseline Score (Time 3) |.
Score Accuracy at Time 2 = 5 – | Time 2 Score – Adjusted Baseline Score (Time 3) |.

In the above formulas, the absolute value parts measure score inaccuracy, that is, the distance between adjusted baseline scores (T3) and T1/T2 scores. The larger the absolute value, the less accurate the score. The absolute value is necessary because there were a sizable percentage of participants whose baseline scores (T3) were higher than their T1/T2 scores. To ease interpretation, we subtracted this absolute value from 5. As a result, the interpretation of score accuracy is straightforward: The larger the value, the higher the score accuracy.

Note that Ellingson et al.’s (2012) approach to analyzing accuracy is slightly different from ours, in that they defined, for instance, Time 1 Accuracy score = Baseline score – Time 1 score. Ellingson et al.’s accuracy scores correctly indicated the degree of accuracy only when baseline scores were lower than T1/T2 scores, that is, when accuracy scores were negative, larger the value (less negative the value), the smaller the distance between T1 baseline scores and T1/T2 scores, the higher the accuracy. However, when accuracy scores were positive, they indicated inaccuracy rather than accuracy, that is, the larger the positive accuracy score, the more distance between baseline scores and T1/T2 scores, the less the accuracy. So, if we had used
Ellingson et al.’s accuracy score approach, it might have created confusion. In contrast, our approach to analyzing accuracy eliminates this confusion.6

**Measures**

In measuring reliabilities, we noticed some of our measures do not meet the assumptions of Cronbach’s alpha. For instance, the emotion measures (i.e., guilt, fear, anger) are not normally distributed. To obtain robust reliability coefficients, we calculated omega (total) in addition to the traditional alpha. Omega (total) is conceptually similar to Cronbach’s alpha and “assesses reliability via a ratio of the variability explained by items compared with the total variance of the entire scale” (McNeish, 2018; p. 416). Omega (total) is appropriate where the strict assumptions of alpha are not satisfied (Bentler, 2007; Geldhof, Preacher, & Zyphur, 2014).

**Personality test.** The Chinese version of a Big Five inventory, the Work Behavior Inventory (WBI; Page, 2011) was used.7 We randomly divided items within each of the Big Five dimension from the WBI into equal halves to create the parallel forms (i.e., Form A and Form B). Each form (107 items) included 32 items measuring Conscientiousness, 20 items measuring Extroversion, 15 items measuring Agreeableness, 20 items measuring Openness, and 20 items measuring Emotional stability. All items were rated on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). One sample item of the Conscientiousness scale was, “I plan my work carefully before I start.” The Omega (total) were .91 (initial block), .93 (main block), and .85 (baseline), respectively in the sample. The Alpha Coefficients were the same as the Omega (total). One sample item of the Extroversion scale was, “I find it easy to start conversations with strangers.” The Omega (total) were .89 (main block) and .83 (baseline),

---

6 Interested readers are referred to the online supplemental materials for a more comprehensive explanation on how our approach to analyzing score accuracy is different from Ellingson et al.’s (2012) approach.
7 The WBI (including the Chinese version) is a proprietary personality measure and its developer, Dr. Ronald C. Page, has granted the authors the permission to use it in the current research.
respectively in the sample. The Alpha Coefficients were the same as the Omega (total). One sample item of the Agreeableness scale was, “I often volunteer to help unfortunate or needy people.” The Omega (total) were .87 (main block) and .75 (baseline), respectively in the sample. The Alpha Coefficients were the same as the Omega (total). One sample item of the Openness scale was, “I frequently combine unrelated concepts in unique ways.” The Omega (total) were .90 (main block) and .86 (baseline), respectively. The Alpha Coefficients were the same as the Omega (total). Finally, one sample item for the Emotional stability scale was, “I worry a lot.” The Omega (total) were .91 (main block) and .84 (baseline), respectively. The Alpha Coefficients were the same as the Omega (total).

**Faking measures.** Following prior studies (e.g., Fan et al. 2012; Lopez, Hou, & Fan, 2019), we used two faking measures. The first faking measure was the 12-item IM scale from the WBI (Page, 2011). One sample item was, “I never tell a lie.” Items were rated on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The Omega (total) were .84 (initial block) and .89 (main block), respectively. The Alpha Coefficients were the same as the Omega (total). Participants whose IM scores were two SDs above the Chinese normative mean were flagged as suspected fakers. This cutoff, which is more stringent than that used by Ellingson et al. (2012), who chose the median IM scores based on a pilot study, was set to reduce the false positives in the field setting. It turned out 340 (23.5%) participants in the final sample met this faking criterion. The second faking measure was the BS scale developed by Fan et al. (2012). This measure, framed as a school activity survey, contained 13 genuine items and 2 bogus items. Respondents were asked to indicate how often they had engaged in various activities during the last 12 months on a 4-point scale (1 = *never*; 4 = *often*). One sample bogus statement was: “Using the Murray-web to locate unpublished research articles.” Responses
endorsing 1 were assigned 0 point, all other responses were assigned 1 point. The Omega (total) were .54 (initial block) and .49 (main block), respectively. The correlation between IM and BS scores was .23 ($p < .05$), demonstrating a modest level of convergent validity, which is similar to what other scholars have reported (e.g., Dunlop et al., 2019). Following Fan et al. (2012), participants whose Time 1 BS scores were 2 points were flagged as fakers, and 120 (8.3%) participants in the final sample met this criterion. Together, 428 (29.6%) participants whose IM scores, or BS scores, or both exceeded the faking criteria were flagged as suspected fakers.

We are aware that scholars have expressed skepticism toward the validity of social desirability (SD) scales (e.g., the IM scale) as faking measures (e.g., Griffith & Peterson, 2008). Such skepticism stemmed primarily from the meta-analytic findings that partialing out SD scores (IM scores) does not affect criterion-related validity of personality scores (e.g., Li & Bragger, 2006; Ones, Viswesvaran, & Reiss, 1996). However, during the last several years, scholars have begun to examine the moderation effect of SD (or IM) scores, moving beyond the traditional suppression effect. Interestingly, several empirical field studies found the hypothesized moderating effect, that is, personality score validity is higher when SD scores are low than when SD scores are high (e.g., Lanyon, Goodstein, & Wershba, 2014; O'Connell, Kung, & Tristan, 2011). Thus, in light of the emerging empirical evidence, we contend that it is probably premature to discard the IM scale as a faking measure. We also note that although the IM and BS scales are not perfect faking measures, they fit with the mid-test warnings context, which requires deciding the faker status after the initial block, and were successfully used in previous mid-test warnings studies (e.g., Ellingson et al., 2012; Fan et al., 2012; Lopez et al., 2019).

**Emotions.** An emotions survey was assembled to measure the three emotions of interest. Items were originally in English. We used the back-translation technique (c.f., Brislin, 1993) to
translate emotion items into Chinese. Specifically, two I/O researchers who are native Chinese
speakers working in the U.S. first translated the English items into Chinese, and then two
American students studying in a Chinese university back-translated the Chinese items into
English. Next, these four individuals held several meetings (through Zoom) during which the
original and back-translated English items were compared, item modifications were made, and
disagreements were resolved. Items were all adjectives, and participants rated the extent to which
they had experienced these emotions after receiving the mid-test warnings or the control message
on a 5-point scale (1 = very slightly or not at all; 5 = extremely). The 6-item guilt scale
developed by Harder and Zalma (1990) was used to measure guilt. Following Ellingson et al.
(2012), the 6 items were converted into 6 adjectives (e.g., guilty, regretful) and were then mixed
with PANAS-X items (Watson & Clark, 1994). The Omega (total) and the Alpha Coefficients
were both .94. The fear scale (6 adjectives; e.g., afraid, frightened) included in PANAS-X
measured fear. The Omega (total) and the Alpha Coefficients were both .89. Finally, the anger
scale (6 adjectives; e.g., angry, irritable) included in PANAS-X was used to measure anger. The
Omega (total) and the Alpha Coefficients were both .91. A series of confirmatory factor analysis
supported the existence of three negative emotion constructs rather than a general construct.8

Perceived test fairness. Perceived fairness of the testing system was measured using the
10-item Chinese questionnaire by Fan et al. (2012), who translated and adapted Tonidandel,
Quinones, and Adams’ (2002) 8-item measure along with 2 items from Smither, Reilly, Millsap,
Pearlman, and Stoffey’s (1993) measure into Chinese. Items were rated on a 5-point Likert scale

8 Details of these confirmatory factor analysis results are included in the online supplemental materials.
ranging from 1 (strongly disagree) to 7 (strongly agree). One sample item is “Overall, I believe the testing system was fair.” The Omega (total) and the Alpha Coefficients were both .82. 9

**Control variables.** Due to participant attrition from T1/T2 to T3, we were concerned that the distribution of the warnings vs. control message might no longer be random. Guided by Bernerth and Aguinis’ (2016) principles of choosing control variables, we chose to include sex and whether participants’ undergraduate degree was obtained from the same college (i.e., college) in our model testing. Sex was chosen because previous studies have documented the associations between sex and emotional experience (e.g., Fischer, Mosquera, van Vianen, & Manstead, 2004). Further, in the current sample, sex was significantly related to anger ($r = -.11, p < .01$). College was chosen, because in the current sample, it was significantly related to Conscientiousness score accuracy at T1 ($r = .11, p < .01$), fear ($r = -.06, p < .05$), and perceived test fairness ($r = -.06, p < .05$).

**Analytic Strategy**

To test our research hypotheses, we conducted a series of multi-group (fakers vs. non-fakers) path analyses using Mplus 8 (Muthén & Muthén, 1998-2017), with guilt, fear, and anger included as three mediators simultaneously. We dichotomized applicants into fakers and non-fakers based on their IM or BS scores, following Fan et al.’s (2012) practice. Such a dichotomization is unavoidable, given that (a) we used more than one faking measure to identify fakers, and (b) in practice the testing system must decide the faker status of every test-takers in order to deliver the warning or control message after the initial block. To model changes in personality score accuracy, we included T1 personality score accuracy in the model and specified a link from T1 score accuracy to T2 score accuracy.

---

9 We have included items for all measures (except for the proprietary WBI items including IM items) in the online supplemental materials.
Preliminary analyses indicated that the three negative emotion scores were skewed and demonstrated floor effects (i.e., at least 15% of participants endorsing the lowest score on the rating scale; c.f., McHorney & Tarlov, 1995). As a result, regular multi-group path analysis with the maximum likelihood (ML) estimation method is no longer appropriate, because it may lead to biases in both parameter estimates and standard errors (Zhu & Gonzalez, 2017). Tobit regression is usually recommended to address floor effects, which treats floor variables as left-censored variables (c.f., Tobin, 1958). However, with censored variables in the path analysis model, it is impossible to calculate means, variances, and covariances as sufficient statistics, and thus commonly used model fit indices cannot be estimated, except for the Chi-square value. Muthén (2009) recommended using the robust ML (MLR) method and working with likelihood-ratio Chi-square testing of nested neighboring models to test if a specific restriction upholds or not; a practice we adopted. The Satorra-Bentler Scaled Chi-square test was used (Satorra, 2000).

Following the typical procedure of conducting multi-group path analyses, we first tested a baseline model in which all path coefficients were freely estimated across the faker and non-faker groups. Next, we tested a constrained model in which all path coefficients except for those associated with control variables were constrained to be equal across the two groups. A significant decrease in model fit from the baseline to constrained model prompted us to identify specific group-variant paths. More specifically, we constrained one path to be equal at a time across the two groups. If the equality constraint did not lead to a significant decrease in model fit (through a Wald Chi-square test), it was retained, and next path constraint was then imposed. If the equality constraint led to a significant decrease in model fit, it was loosened, and a next path constraint was imposed. This procedure continued until all hypothesized paths were tested.\(^\text{10}\)

\(^{10}\) MPlus code for multi-group path analyses are included in the online supplemental materials.
We used the bias-corrected bootstrapping approach based on 5,000 resamples (Preacher & Hayes, 2008) to test the indirect effects for the faker and non-faker groups. Note that the bootstrapping option is not available for the MLR method in Mplus. Following Muthén’s recommendation, we used the MLR method to obtain model fits, but the ML method to obtain bootstrapping-based confidence intervals for indirect effects. As noted by Muthén, the MLR and ML methods yield different standard errors, but exactly same parameter estimates. As such, bootstrapping-based confidence intervals should not differ between the two methods.

Results

Means and standard deviations of personality scores, score accuracy, emotions, and perceived test fairness across time for the four cells are presented in Table 1. Cohen’s $d$s indicating effect sizes of the treatment among fakers and non-fakers are also reported in Table 1. Correlations among study variables for the entire sample are presented in Table 2. As can be seen, the mid-test warnings increased score accuracy, but much more so for fakers ($\Delta A = .24$) than for non-fakers ($\Delta A = .08$); the control message had minimal influence on changes in score accuracy for fakers ($\Delta A = .04$) and non-fakers ($\Delta A = .03$). Note that T2 personality score accuracy was quite similar across the Big Five traits within each of the four cells. These results support the carry-over effect of the mid-test warnings reported by Fan et al. (2012), that is, the warning effect generalized to personality scales not included in the initial block. However, our primary focus here was on Conscientiousness scores.

Multi-group Path Analysis Results

The multi-group baseline model yielded a Log Likelihood value of -7141.83, with $df = 60$, whereas the constrained model yielded a Log Likelihood value of -7169.157, with $df = 48$. A Satorra-Bentler scaled $\chi^2$ difference test was significant: $\Delta \chi^2 (\Delta df = 12) = 45.03, p < .01$. This
suggests that at least some of the path coefficients were different across the faker and non-faker groups. We then proceeded to examine the equality of path coefficients, one at a time, across the two treatment groups. Figures 3 depicts the finalized multi-group path analysis model.

H1 predicted that applicants receiving the mid-test warnings should report a higher level of guilt than those receiving the control message, and that the effect should be stronger for fakers than for non-fakers. The Wald Chi-square test indicates that imposing an equality constraint on the treatment – guilt path led to significantly worse model fit: Wald $\chi^2 (df = 1) = 5.20, p < .05$. Figure 3 shows that the above path was stronger for fakers ($B = 1.26, se = .11, p < .01$) than for non-fakers ($B = .96, se = .07, p < .01$), with both being significant. Thus, H1 was supported.

H2 predicted that applicants receiving the mid-test warnings should report a higher level of fear than those receiving the control message, with no moderation effect of faker status. Consistent with our expectation, the Wald Chi-square test indicates that imposing an equality constraint on the treatment – fear path did not lead to significantly worse model fit: Wald $\chi^2 (df = 1) = .63, p = .43$. Figure 3 shows the above path was positive and significant for both fakers and non-fakers ($B = .67, se = .05, p < .01$), thus supporting H2.

H3 predicted that the treatment effect on anger should be positive, and that the effect should be stronger for non-fakers than for fakers. The Wald Chi-square test indicates that imposing an equality constraint on the treatment – anger path did not lead to significantly worse model fit: Wald $\chi^2 (df = 1) = .84, p = .36$. Figure 3 shows that the path was positive for both fakers and non-fakers ($B = .59, se = .05, p < .01$). Thus, H3 was partially supported.

H4a predicted a stronger positive association between guilt and change in score accuracy for fakers than for non-fakers. The Wald Chi-square test reveals that imposing an equality constraint on the guilt – change in score accuracy did not lead to significantly worse model fit:
Wald $\chi^2 (df = 1) = .99, p = .32$. Figure 3 shows that the above path was positive and significant for both fakers and non-fakers ($B = .03, se = .01, p < .05$). Thus, H4a was not supported.

Hypothesis 4b predicted that faker status should moderate the indirect link of treatment – guilt – change in score accuracy. The Wald Chi-square test indicates that imposing an equality constraint on the indirect link did not lead to significantly worse model fit: Wald $\chi^2 (df = 1) = 2.65, p = .10$. As such, H4b was not supported. Nevertheless, bootstrapping-based CIs reveal (see Table 3) that the indirect link was significant for fakers (95% CI [.01, .13]), but not significant for non-fakers (95% CI [–.01, .06]), indicating that guilt mediated the treatment effect on an increase in personality score accuracy for fakers, but not for non-fakers.

H5a proposed a positive association between fear and change in score accuracy for fakers, but a negative association for non-fakers. The Wald Chi-square test reveals that imposing an equality constraint on the fear – change in score accuracy link led to significantly worse model fit: Wald $\chi^2 (df = 1) = 4.64, p < .05$. Figure 3 shows that the path was positive, yet not significant for fakers ($B = .01, se = .02, p = .50$), but was significant and negative for non-fakers ($B = -.05, se = .02, p < .01$). Thus, H5a received partial support. Another Wald Chi-square test reveals that the moderated mediation effect was significant, in that imposing an equality constraint on the treatment – fear – change in score accuracy indirect link led to significantly worse model fit: Wald $\chi^2 (df = 1) = 4.55, p < .05$. According to bootstrapping-based CIs (see Table 3), the indirect link was not significant for fakers (95% CI [–.04, .04]), but was significantly negative for non-fakers (95% CI [–.05, –.01]). Thus, H5b was partially supported.

H6a proposed that the relationship between anger and change in score accuracy would be more negative for non-fakers than for fakers. Inconsistent with our expectation, the Wald Chi-square test indicates that imposing an equality constraint on the anger – change in score accuracy
link did not lead to a significantly worse model fit: Wald $\chi^2 (df = 1) = .01, p = .92$. Figure 3 shows that the link was non-significant for both fakers and non-fakers ($B = -.02, se = .02, p = .26$). H6a was not supported. The moderated mediation effect (H6b) was not supported, either, because both the first-stage (H3) and second-stage (H6a) moderations were not supported.

For exploratory purposes, we tested whether the direct link between treatment and change in score accuracy differed across the two groups. The Wald Chi-square test indicates that imposing an equality constraint on the treatment – change in score accuracy link led to significantly worse model fit: Wald $\chi^2 (df = 1) = 15.51, p < .01$. Figure 3 shows that the path was not significant for non-fakers ($B = .02, se = .02, p = .16$), but was significant for fakers ($B = .16, se = .03, p < .01$).

Research questions (RQ) 1-3 concern the effects of the mid-test warnings on applicants’ perceived test fairness. Regarding the relationship between guilt and perceived test fairness, the Wald Chi-square test reveals that imposing an equality constraint on the guilt – perceived test fairness link did not lead to a significantly worse model fit: Wald $\chi^2 (df = 1) = 1.84, p = .17$. Figure 3 shows this path was non-significant for both fakers and non-fakers ($B = .04, se = .03, p = .10$). Another Wald Chi-square test indicates that imposing an equality constraint on the indirect link of treatment – guilt – perceived test fairness did not lead to a significantly worse model fit: Wald $\chi^2 (df = 1) = 1.86, p = .17$. Nevertheless, bootstrapping-based CIs reveal (see Table 3) that the indirect effect of treatment – guilt – perceived test fairness was not significant for fakers (95% CI [–.12, .11]), but was significantly positive for non-fakers (95% CI [.01, .12]). Therefore, guilt did not mediate the treatment effect on perceived test fairness for fakers, but mediated the treatment effect on perceived test fairness for non-fakers.
As for the link between fear and perceived test fairness, the Wald Chi-square test indicates that imposing an equality constraint on the fear – perceived test fairness link did not lead to a significantly worse model fit: Wald $\chi^2 (df = 1) = 1.64, p = .20$. This path was non-significant for both fakers and non-fakers ($B = -.02, se = .03, p = .61$). The mediation analyses results based on bootstrapping CIs show (see Table 3) that the indirect link of treatment – fear – perceived test fairness was not significant for both fakers (95% CI [–.11, .06]) and non-fakers (95% CI [–.05, .04]). Therefore, fear was not related to perceived test fairness, and fear did not mediate the treatment – perceived test fairness link. Further, as both the first-stage and second-stage moderations were not significant, faker status did not moderate the indirect link of treatment – fear – perceived test fairness.

When imposing an equality constraint on the anger – perceived test fairness link, the model fit did not become significantly worse: Wald $\chi^2 (df = 1) = .05, p = .83$. This path was negative and significant for both fakers and non-fakers ($B = –.23, se = .03, p < .01$). According to Table 3, bootstrapping-based CIs indicate that the indirect link of treatment – anger – perceived test fairness was significantly negative for both fakers (95% CI [–.19, –.03]) and non-fakers (95% CI [–.21, –.11]). Therefore, anger was negatively related to perceived test fairness, and anger mediated the treatment – perceived test fairness link; however, faker status did not moderate the indirect effect of treatment – anger – perceived test fairness, because according to earlier analyses, both the first-stage and second-stage moderations were not significant.

We also tested whether the direct link between treatment and perceived test fairness differed across fakers and non-fakers. The Wald Chi-square test indicates that imposing an equality constraint on the treatment – perceived test fairness link did not lead to a significantly
worse model fit: Wald $\chi^2 (df = 1) = 2.30, p = .13$. This direct link was significant for both fakers and non-fakers ($B = -.20, se = .04, p < .01$).

Finally, we conducted two separate sets of additional analyses to examine whether findings were robust to whether IM and BS scores were treated as dichotomized or continuous (or ordinal) scores. In the first set of additional analyses, we compared the results based on using IM scores as the lone dichotomized vs. continuous moderator. It turned out these alternative models yielded similar results and identical statistical conclusions. In the second set of additional analyses, we compared the results based on using BS scores as the lone dichotomized vs. ordinal moderator. These alternative models yielded similar results and identical statistical conclusions. Thus, we conclude that dichotomizing IM and BS scores did not unduly affect our main findings.

Discussion

Applicant faking has been one of most recurring objections against the applied use of personality scales in personnel selection throughout the history of personality (Hough & Oswald, 2008; Sackett et al., 2017). Despite the voluminous research devoted to this topic, the role of emotions has been severely under-studied. Drawing from an affect-based theoretical framework (Nabi, 1999) that enabled us to analyze multiple emotions in a systematic way, the present study investigated how several discrete negative emotions (i.e., guilt, fear, and anger) triggered by mid-test warnings influenced applicant faking and perceived test fairness. We also examined faker status as a boundary condition for the proposed mechanisms. An ambitious and rigorous within-subjects field experiment was conducted with a large sample of graduate school applicants in China to test our hypotheses and explore several research questions.

---

11 Detailed results of alternative modeling are included in the online supplemental materials.
Our overall conclusion is that guilt, fear, and anger each played a unique role in our conceptual model. That is, guilt explained how mid-test warnings helped boost personality score accuracy among fakers, whereas fear explained why non-fakers over-corrected their personality scores after being warned. Anger also mattered, in that it accounted for why mid-test warnings reduced perceived test fairness.

**Theoretical Implications**

The present study makes at least three contributions to applicant faking and selection research. First, the present study is one of the first to empirically demonstrate that a variety of negative emotions play a role in applicant faking. We found that *guilt* acted as a key mechanism for warnings’ effect on score accuracy increase among fakers. In other words, mid-test warnings evoked a modest level of guilt in fakers, and that guilt motivated them to remedy the relationship with the test administrator by providing more honest responses in the main block. We extend previous finding that guilt predicted score accuracy increase under the warning condition (Ellingson et al., 2012) by substantiating the direct link between warning and guilt, and establishing the mediating role of guilt in the relationship between warnings and score accuracy increase among fakers. Although mid-test warnings evoked a somewhat weaker level of guilt in non-fakers, which in turn was associated with an increase in personality score accuracy, the indirect link of treatment – guilt – increase in score accuracy was not significant for non-fakers.

Importantly, this study also sheds light on the underlying mechanism behind overcorrecting personality scores upon receiving a warning. That is, *fear* turned out to be the key mechanism of the effect of mid-test warnings on a decrease in personality score accuracy among non-fakers. This finding suggests that when a modest level of fear is evoked by the mid-test warnings, although it had little impact on fakers, it prompted non-fakers to underreport their
personality scores to avoid potential punishment. To our knowledge, the current study is the first to illuminate this mechanism underlying score over-correction due to mid-test warnings. One important theoretical implication of our results related to guilt and fear is that current faking models (e.g., Goffin & Boyd, 2009; Shoss & Strube, 2011) should be revised and broadened to incorporate discrete negative emotions. Accordingly, these models might more fully account for applicant faking processes.

Second, this study extended a small body of emerging research on mid-test warnings (e.g., Ellingson et al., 2012; Fan et al., 2012) by examining its emotional mechanisms and testing faker status as a boundary condition. Results revealed that the mid-test warnings exerted influence on fakers and non-fakers through different emotional pathways, thus advancing our understanding of why and for whom mid-test warnings work.

Third, the present study shifts the attention from traditional cognitive pathways (e.g., Ployhart & Harold, 2004) through which applicant perceptions are formed to emotional pathways. Importantly, we add to knowledge in this area by discovering that anger is the emotion that serves as the main reason why mid-test warnings may harm applicants’ perceived test fairness for both fakers and non-fakers. Given that we included all three negative emotions simultaneously in model testing, this finding suggests that it is not the general negative emotions, but the specific emotion of anger that led to lowered perceived test fairness. This can be explained by previous research showing that anger elicits an approach action tendency to correct the transgressor and exact revenge (Carver & Harmon-Jones, 2009; Gibson & Callister, 2010; Leach, 2008). Interestingly, we also found a positive indirect link of treatment – guilt – perceived test fairness for non-fakers, but not for fakers. One possible explanation is that by lowering their inflated scores, warned fakers’ guilt was largely dissipated. However, warned non-fakers did not
have much room for lowering their scores; instead, they resorted to providing positive evaluations of the test system as a way to restore the relationship with the test administrator. In any event, as one implication, current theorizing on applicant perceptions should be broadened to account for the role of emotions.

**Study Limitations**

As noted in the review process, the first limitation is that although applicants who received mid-test warnings reported significantly higher levels of negative emotions than those receiving the control message, the warned applicants on average reported having experienced only a little negative emotion (2 on a 5-point scale ranging from 1 = *very slightly or not at all* to 5 = *extremely*). This begs the question of whether the warned applicants actually experienced much negative emotions. We acknowledge the findings as they are, but would like to note that our findings were within the normal range when viewed within the larger emotion research context. In the emotions literature, there is an interesting asymmetry that individuals tend to report much lower levels of negative emotions than positive emotions, and these result patterns have been observed consistently across a wide variety of study designs (e.g., Glasø, Vie, Holmdal, & Einarsen, 2011; Liu, Song, Li, & Liao, 2017). 12 One possible explanation is that individuals preferentially suppress their expression of negative emotions compared to positive emotions, because negative emotional displays are socially inappropriate (Ekman & Frisen, 1969). In any event, the result patterns observed in the larger emotion literature are similar to what we found, suggesting that our findings were not anomaly.

As another potential limitation, one might argue that our findings are an artifact of our research design because we did not measure emotions immediately after the middle messages,

---

12 Detailed information about these empirical studies on emotions can be found in the online supplemental materials.
but rather waited until the entire test was completed. We conducted another mid-test warnings study in the following year in the same school and the same selection context, but with different purposes and a slightly different research design. In that study, we measured several negative and positive emotions before the warning, shortly after the warning, in the middle and then toward the end of the personality test. Very similar result patterns emerged; that is, applicants reported much lower negative emotion mean scores than positive emotion mean scores even when emotions were measured shortly after the warning. Therefore, the low scores in negative emotions in the present study did not appear to be caused by the timing of emotion measurement. Nevertheless, future research should consider measuring emotions via physiological responses (e.g., cardiovascular activity, electrodermal responding, and somatic activity; Ax, 1953), facial expression evaluations (e.g., Bailenson et al., 2008), and wearable sensors (e.g., Matusik, Heidl, Hollenbeck, Yu, Lee, & Howe, 2018).

Third, we observed small effect sizes for the path coefficients from emotions to personality score accuracy change and for the mediation links. Despite many significant paths, they might be attributed to the large sample size of the current study. Partially alleviating this concern was the aforementioned observation that guilt, fear, and anger, when tested simultaneously in the path analysis model, exhibited unique effects, which cannot be explained by the large sample size alone. The small effect sizes also suggest that emotional mechanisms might be only one piece in the faking puzzle, and other factors are also relevant. For instance, cognitive factors such as the perception of being watched (e.g., Bateson, Nettle, & Roberts, 2006) and perceived risks (Slovic, 1987) may provide additional explanations of the warning effect. Therefore, we call for future research to empirically examine affective and cognitive mechanisms in tandem.

---

13 More detailed information about this study and findings on emotions can be found in the online supplemental materials.
The fourth limitation is that personality scores of unwarned non-fakers also decreased non-trivially from the main block (T2) to the baseline (T3), suggesting that there might be mild fakers in the non-faker group. This implies that our faking measures along with stringent cutoff scores might have captured only extreme fakers, potentially categorizing mild fakers into the non-faker group. As a result, our results associated with the non-faker group should be interpreted with caution. For example, we found a significant treatment effect on guilt among non-fakers, which is inconsistent with the theoretical expectation, but which can be explained by the non-faker group consisting of both real non-fakers and mild fakers. We speculate that with faking measures capable of identifying a wider variety of fakers, our results might be more clear-cut for non-fakers. Unfortunately, to our best knowledge, there currently does not exist faking measures that may accurately identify all types of fakers. Our faking measures, albeit not perfect, are suited to the mid-test warning contexts, were used successfully in previous similar studies (e.g., Ellingson et al., 2012; Fan et al., 2012), and provided more conservative tests of our research hypotheses.

Finally, we note some caveats in terms of external validity. For example, given the relatively high selection ratio at the campus interview stage (around 70%), we do not know whether our results might generalize to high stakes settings with a more stringent selection ratio. Further, given the Chinese context, it is unclear whether our findings might generalize to Western societies. Whereas Chinese societies are considered a typical collectivistic culture that emphasizes interdependence and following the order of authority, Western societies such as America are considered individualistic cultures that value independence and pursuing personal interests (Markus & Kitayama, 1991). It is conceivable that compared with Chinese applicants, Western applicants might be less willing to heed mid-test warnings, rendering them less effective in boosting personality score accuracy, but also less likely resulting in over-correcting
personality scores. Future research is thus needed to replicate our findings in such high stakes situations and in different cultures.

**Future Research Directions**

We contend that an affective perspective represents an important and useful new angle to selection and faking research. As one example, future research should examine the role of emotion in the faking processes in general. For instance, McFarland and Ryan (2006) theorized that subjective norm may influence faking such that applicants who perceive faking is socially accepted should be more likely to engage in faking. We speculate that potential mechanisms for such an effect might be due to applicants experiencing less guilt and fear. As another example, future studies are needed to investigate the effects of mid-test warnings with different message content and the role of emotions underlying their effects. As of now, the majority of mid-test warnings have used the traditional detection/accusation and consequence warnings, which have a strong distrusting and threatening tone. Interestingly, scholars have suggested several new, friendlier warnings (e.g., Burns et al., 2015, Pace & Borman, 2006, Turcu, 2011). Yet, we are still not clear regarding their relative effects and underlying mechanisms. In addition, future researchers may examine the role of emotions in other selection methods such as interviews and assessment centers. For instance, there is evidence that interviewees’ use of impression management tactics predicted interview scores (e.g., Swider, Barrick, Harris, & Stoverink, 2011). These tactics might evoke various emotions in interviewers, who might rely on these affects as information in deciding interview scores.

Finally, future research should examine whether individual differences might moderate the effects of mid-test warnings. For instance, psychopathy at workplace has received increasing attention (e.g., Spain, Harms, & LeBreton, 2014; Stevens, Deuling, & Armenakis, 2012).
Individuals high in psychopathy are characterized as being manipulative, callous, and lacking empathy, and guilt (Barelds, Wisse, Sanders, & Laurijssen, 2018). Psychopathy has been linked to self-serving and unethical behaviors (Smith & Lilienfeld, 2013; Williams, 2014). We speculate that applicants with high levels of psychopathy are more likely to fake on selection procedures because it is in favor of their personal benefit, and they are less likely to feel guilt or fear, but more likely to feel anger after receiving the warnings due to their defiant nature. Future research should examine whether psychopathy weakens the effects of mid-test warnings in boosting personality score accuracy, but worsens the effects on applicant perception.

**Practical Implications**

Findings of the present study have several important practical implications. First of all, we would like to emphasize that in field settings the mid-test warnings should only be delivered to suspected fakers, but not to non-fakers. This is because our findings suggest that non-faker’s personality score accuracy would decrease and negative applicant perceptions would result, if they received warnings. The true experimental design (with the random assignment of warning vs. control messages) used in the current study was for research purposes, and should not be used in real-world selection contexts.

As another implication, this study suggests that the content and style of warning messages could be improved. To promote the increase of personality score accuracy, organizations interested in using mid-test warnings to address applicant faking should strive to make sure that the warnings evoke feelings of guilt, while minimizing feelings of fear and anger. We recommend that when developing mid-test warning messages, organizations pilot test several versions of mid-test warnings and use subject matter experts to make sure they yield the desired
types of emotions. On the basis of these pilot tests, organizations should then consider modifying various components of the warning messages to achieve desired effects.

As still another implication, organizations should be aware that organizational initiatives could trigger emotions in applicants, which in turn, could influence their test performance and reactions toward the selection process. For instance, hiring organizations should strive to establish rapport and trust with prospective applicants such that applicants may experience guilt when trying to fake on selection tests and engaged in other types of dishonest behaviors. Similarly, by thoroughly explaining the selection processes to job applicants, hiring organizations may potentially reduce anger among applicants, which in turn might lead to more cooperative behaviors by applicants. However, there might be a trade-off between the effectiveness and friendliness of the initiatives. Practitioners might have to strike a balance between the two aspects when applying a warning message in real selection contexts. In sum, we are optimistic that with programmatic research, we eventually will be able to provide selection professionals with well-developed mid-test warning strategies that may be used to effectively manage applicant faking in real-world selection settings.

Reference


Table 1.

**Means and Standard Deviations of Personality Scores, Accuracy, Emotions, and Perceived Test Fairness across Study Cells**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
<td>Acc T1</td>
</tr>
<tr>
<td>C</td>
<td>4.22 (.36)</td>
<td>4.16 (.39)</td>
<td>3.68 (.38)</td>
<td>4.43 (.37)</td>
</tr>
<tr>
<td>E</td>
<td>3.63 (.44)</td>
<td>3.02 (.48)</td>
<td>4.36 (.44)</td>
<td>3.14 (.48)</td>
</tr>
<tr>
<td>O</td>
<td>3.76 (.41)</td>
<td>3.17 (.49)</td>
<td>4.35 (.43)</td>
<td>3.14 (.48)</td>
</tr>
<tr>
<td>A</td>
<td>3.97 (.41)</td>
<td>3.59 (.38)</td>
<td>4.52 (.34)</td>
<td>3.39 (.46)</td>
</tr>
<tr>
<td>ET</td>
<td>3.62 (.47)</td>
<td>3.23 (.51)</td>
<td>4.49 (.41)</td>
<td>2.90 (.46)</td>
</tr>
<tr>
<td>Guilt</td>
<td>1.33 (.53)</td>
<td>2.13 (.95)</td>
<td>1.99 (.86)</td>
<td>1.99 (.86)</td>
</tr>
<tr>
<td>Fear</td>
<td>1.41 (.55)</td>
<td>1.99 (.86)</td>
<td>1.99 (.86)</td>
<td>1.99 (.86)</td>
</tr>
<tr>
<td>Anger</td>
<td>1.30 (.50)</td>
<td>1.58 (.62)</td>
<td>1.58 (.62)</td>
<td>1.58 (.62)</td>
</tr>
<tr>
<td>Fairness</td>
<td>3.78 (.70)</td>
<td>3.47 (.62)</td>
<td>3.47 (.62)</td>
<td>3.47 (.62)</td>
</tr>
</tbody>
</table>

**Note.** C = Conscientiousness scale. E = Extroversion scale. O = Openness to Experience scale. A = Agreeableness scale. ET = Emotional Stability scale. T1 = Time 1 (initial block). T2 = Time 2 (main block). T3 = Time 3 (baseline). Acc_T1 = score accuracy at Time 1. Acc_T2 = score accuracy at Time 2. ΔA = A_T2 – A_T1. Cohen’s ds for the treatment effect (warning vs. control message) are calculated based on personality scores at the main block (T2), emotion scores, and perceived test fairness scores among fakers and non-fakers, respectively.
### Table 2.
Means, Standard Deviations, Reliabilities, and Correlations among Study Variables

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sex</td>
<td>1.69</td>
<td>.46</td>
<td>(—)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. College</td>
<td>.34</td>
<td>.47</td>
<td>-.01</td>
<td>(—)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Warning</td>
<td>.50</td>
<td>.50</td>
<td>.02</td>
<td>-.05</td>
<td>(—)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. IM_T1</td>
<td>40.24</td>
<td>7.23</td>
<td>.16**</td>
<td>-.11**</td>
<td>.03</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. BS_T1</td>
<td>.50</td>
<td>.69</td>
<td>.02</td>
<td>.03</td>
<td>.03</td>
<td>.23**</td>
<td>(.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. IM_FakerT1</td>
<td>.23</td>
<td>.42</td>
<td>.09**</td>
<td>-.10**</td>
<td>.03</td>
<td>.73**</td>
<td>.19**</td>
<td>(—)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. BS_FakerT1</td>
<td>.11</td>
<td>.31</td>
<td>-.01</td>
<td>.03</td>
<td>.08**</td>
<td>.19**</td>
<td>.77**</td>
<td>.18**</td>
<td>(—)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Faker_T1</td>
<td>.30</td>
<td>.46</td>
<td>.07**</td>
<td>-.08**</td>
<td>.06</td>
<td>.66**</td>
<td>.47**</td>
<td>.86**</td>
<td>.54**</td>
<td>(—)</td>
<td></td>
</tr>
<tr>
<td>9. C_T1</td>
<td>3.90</td>
<td>.41</td>
<td>.02</td>
<td>-.13**</td>
<td>.04</td>
<td>.68**</td>
<td>.27**</td>
<td>.51**</td>
<td>.20**</td>
<td>.49**</td>
<td>(.91)</td>
</tr>
<tr>
<td>10. IM_T2</td>
<td>36.04</td>
<td>8.12</td>
<td>.16**</td>
<td>-.06*</td>
<td>-.38**</td>
<td>.75**</td>
<td>.14**</td>
<td>.54**</td>
<td>.09**</td>
<td>.46**</td>
<td>.47**</td>
</tr>
<tr>
<td>11. BS_T2</td>
<td>.13</td>
<td>.34</td>
<td>.03</td>
<td>.03</td>
<td>-.11**</td>
<td>.13**</td>
<td>.48**</td>
<td>.12**</td>
<td>.44**</td>
<td>.26**</td>
<td>.13**</td>
</tr>
<tr>
<td>12. C_T2</td>
<td>3.62</td>
<td>.49</td>
<td>.03</td>
<td>-.05*</td>
<td>-.43**</td>
<td>.48**</td>
<td>.19**</td>
<td>.36**</td>
<td>.12**</td>
<td>.34**</td>
<td>.68**</td>
</tr>
<tr>
<td>13. E_T2</td>
<td>3.16</td>
<td>.54</td>
<td>.05</td>
<td>-.04</td>
<td>-.41**</td>
<td>.25**</td>
<td>.16**</td>
<td>.22**</td>
<td>.11**</td>
<td>.21**</td>
<td>.38**</td>
</tr>
<tr>
<td>14. A_T2</td>
<td>3.48</td>
<td>.52</td>
<td>-.02</td>
<td>-.06*</td>
<td>-.49**</td>
<td>.35**</td>
<td>.11**</td>
<td>.26**</td>
<td>.05*</td>
<td>.23**</td>
<td>.42**</td>
</tr>
<tr>
<td>15. O_T2</td>
<td>3.22</td>
<td>.55</td>
<td>-.10**</td>
<td>-.03</td>
<td>-.46**</td>
<td>.31**</td>
<td>.18**</td>
<td>.24**</td>
<td>.11**</td>
<td>.24**</td>
<td>.41**</td>
</tr>
<tr>
<td>16. ET_T2</td>
<td>3.02</td>
<td>.57</td>
<td>-.10**</td>
<td>-.01</td>
<td>-.51**</td>
<td>.38**</td>
<td>.13**</td>
<td>.28**</td>
<td>.07**</td>
<td>.24**</td>
<td>.39**</td>
</tr>
<tr>
<td>17. C_T3</td>
<td>3.53</td>
<td>.40</td>
<td>.11**</td>
<td>.04</td>
<td>-.08**</td>
<td>.30**</td>
<td>.12**</td>
<td>.24**</td>
<td>.07**</td>
<td>.21**</td>
<td>.41**</td>
</tr>
<tr>
<td>18. E_T3</td>
<td>2.95</td>
<td>.48</td>
<td>-.11**</td>
<td>-.02</td>
<td>-.08**</td>
<td>.03</td>
<td>.09**</td>
<td>.06*</td>
<td>.05*</td>
<td>.08**</td>
<td>.18**</td>
</tr>
<tr>
<td>19. A_T3</td>
<td>3.48</td>
<td>.36</td>
<td>-.03</td>
<td>-.11**</td>
<td>.21**</td>
<td>.07**</td>
<td>.18**</td>
<td>.04</td>
<td>.15**</td>
<td>.25**</td>
<td></td>
</tr>
<tr>
<td>20. O_T3</td>
<td>3.12</td>
<td>.49</td>
<td>-.23**</td>
<td>-.02</td>
<td>-.07**</td>
<td>.09**</td>
<td>.07**</td>
<td>.02</td>
<td>.07**</td>
<td>.22**</td>
<td></td>
</tr>
<tr>
<td>21. ET_T3</td>
<td>3.11</td>
<td>.47</td>
<td>-.15**</td>
<td>-.01</td>
<td>-.12**</td>
<td>.15**</td>
<td>.04</td>
<td>.13**</td>
<td>.01</td>
<td>.10**</td>
<td>.22**</td>
</tr>
<tr>
<td>22. AccC_T1</td>
<td>4.54</td>
<td>.35</td>
<td>.08**</td>
<td>.11**</td>
<td>-.10**</td>
<td>-.29**</td>
<td>-.11**</td>
<td>-.23**</td>
<td>-.10**</td>
<td>-.24**</td>
<td>-.46**</td>
</tr>
<tr>
<td>23. AccC_T2</td>
<td>4.62</td>
<td>.30</td>
<td>.05</td>
<td>.03</td>
<td>.06*</td>
<td>-.12**</td>
<td>-.07**</td>
<td>-.11**</td>
<td>-.06*</td>
<td>-.12**</td>
<td>-.20**</td>
</tr>
<tr>
<td>24. AccE_T2</td>
<td>4.60</td>
<td>.33</td>
<td>.01</td>
<td>.04</td>
<td>.22**</td>
<td>-.18**</td>
<td>-.04</td>
<td>-.16**</td>
<td>-.03</td>
<td>-.13**</td>
<td>-.22**</td>
</tr>
<tr>
<td>25. AccA_T2</td>
<td>4.61</td>
<td>.30</td>
<td>.01</td>
<td>.02</td>
<td>-.05</td>
<td>-.03</td>
<td>.01</td>
<td>-.07**</td>
<td>-.01</td>
<td>-.07**</td>
<td>-.07**</td>
</tr>
<tr>
<td>26. AccO_T2</td>
<td>4.59</td>
<td>.33</td>
<td>-.07*</td>
<td>.03</td>
<td>.14**</td>
<td>-.16**</td>
<td>-.06</td>
<td>-.17**</td>
<td>-.07**</td>
<td>-.17**</td>
<td>-.19**</td>
</tr>
<tr>
<td>27. AccET_T2</td>
<td>4.57</td>
<td>.35</td>
<td>-.01</td>
<td>.01</td>
<td>-.12**</td>
<td>-.09**</td>
<td>-.03</td>
<td>-.09**</td>
<td>-.01</td>
<td>-.08**</td>
<td>-.12**</td>
</tr>
<tr>
<td>28. Guilt</td>
<td>1.68</td>
<td>.82</td>
<td>-.03</td>
<td>-.03</td>
<td>.41**</td>
<td>-.01</td>
<td>.06*</td>
<td>.02</td>
<td>.11**</td>
<td>.07**</td>
<td>.02</td>
</tr>
<tr>
<td>29. Fear</td>
<td>1.71</td>
<td>.76</td>
<td>-.02</td>
<td>-.06*</td>
<td>.37**</td>
<td>-.06*</td>
<td>.01</td>
<td>-.02</td>
<td>.05</td>
<td>.01</td>
<td>-.06*</td>
</tr>
<tr>
<td>30. Anger</td>
<td>1.50</td>
<td>.63</td>
<td>-.11**</td>
<td>.05</td>
<td>.29**</td>
<td>-.14**</td>
<td>-.02</td>
<td>-.12**</td>
<td>.05</td>
<td>-.05</td>
<td>-.12**</td>
</tr>
<tr>
<td>31. Test Fairness</td>
<td>3.49</td>
<td>.63</td>
<td>-.06*</td>
<td>-.20**</td>
<td>.18**</td>
<td>.07**</td>
<td>.17**</td>
<td>.04</td>
<td>.12**</td>
<td>.20**</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** *N* = 1447. IM = Impression Management. BS = Bogus Statement. IM_Faker = Faker status based on IM score. BS_Faker = Faker status based on BS score. Faker = Faker status based on IM or BS score. C = Conscientiousness. E = Extroversion. A = Agreeableness. O = Openness. ET = Emotional Stability. T1 = initial block. T2 = main block. T3 = baseline. Acc = score accuracy. Numbers on the diagonal are Omega (total) coefficients for scales. We also estimated the traditional Cronbach’s Alphas, whose values were identical to Omega (total) coefficients, except for BS scores whose Alpha Coefficients were .54 and .49 at T1 and T2, respectively. *p < .05. **p < .01.
Table 2 (Cont’d)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Warning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. IM_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. BS_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. IM_FakerT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. BS_FakerT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Faker_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. C_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. IM_T2</td>
<td>(.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. BS_T2</td>
<td>.17**</td>
<td>(.56)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. C_T2</td>
<td>.66**</td>
<td>.19**</td>
<td>(93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. E_T2</td>
<td>.43**</td>
<td>.17**</td>
<td>.63**</td>
<td>(.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. A_T2</td>
<td>.59**</td>
<td>.16**</td>
<td>.72**</td>
<td>.57**</td>
<td>(.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. O_T2</td>
<td>.52**</td>
<td>.20**</td>
<td>.72**</td>
<td>.69**</td>
<td>.61**</td>
<td>(.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. ET_T2</td>
<td>.62**</td>
<td>.21**</td>
<td>.70**</td>
<td>.63**</td>
<td>.69**</td>
<td>.70**</td>
<td>(.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. C_T3</td>
<td>.29**</td>
<td>.08**</td>
<td>.44**</td>
<td>.21**</td>
<td>.26**</td>
<td>.24**</td>
<td>.26**</td>
<td>(.85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. E_T3</td>
<td>.04</td>
<td>.03</td>
<td>.24**</td>
<td>.58**</td>
<td>.16**</td>
<td>.33**</td>
<td>.23**</td>
<td>.28**</td>
<td>(.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. A_T3</td>
<td>.23**</td>
<td>.08**</td>
<td>.29**</td>
<td>.22**</td>
<td>.40**</td>
<td>.22**</td>
<td>.25**</td>
<td>.44**</td>
<td>.24**</td>
<td>(.75)</td>
<td></td>
</tr>
<tr>
<td>20. O_T3</td>
<td>.09**</td>
<td>.03</td>
<td>.26**</td>
<td>.39**</td>
<td>.17**</td>
<td>.51**</td>
<td>.25**</td>
<td>.28**</td>
<td>.58**</td>
<td>.18**</td>
<td>(.86)</td>
</tr>
<tr>
<td>21. ET_T3</td>
<td>.18**</td>
<td>.06*</td>
<td>.28**</td>
<td>.33**</td>
<td>.26**</td>
<td>.35**</td>
<td>.46**</td>
<td>.39**</td>
<td>.43**</td>
<td>.38**</td>
<td>.48**</td>
</tr>
<tr>
<td>22. AccC_T1</td>
<td>-.13**</td>
<td>-.04</td>
<td>-.18**</td>
<td>-.15**</td>
<td>-.13**</td>
<td>-.13**</td>
<td>-.10**</td>
<td>.47**</td>
<td>.15**</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>23. AccC_T2</td>
<td>-.12**</td>
<td>-.04</td>
<td>-.16**</td>
<td>-.18**</td>
<td>-.12**</td>
<td>-.15**</td>
<td>-.13**</td>
<td>.15**</td>
<td>-.03</td>
<td>.03</td>
<td>-.01</td>
</tr>
<tr>
<td>24. AccE_T2</td>
<td>-.28**</td>
<td>-.10**</td>
<td>-.31**</td>
<td>-.30**</td>
<td>-.28**</td>
<td>-.28**</td>
<td>-.27**</td>
<td>.03</td>
<td>.18**</td>
<td>-.02</td>
<td>.07**</td>
</tr>
<tr>
<td>25. AccA_T2</td>
<td>.02</td>
<td>.01</td>
<td>.04</td>
<td>.02</td>
<td>.13**</td>
<td>.03</td>
<td>.06*</td>
<td>.05</td>
<td>.03</td>
<td>-.09**</td>
<td>-.01</td>
</tr>
<tr>
<td>26. AccO_T2</td>
<td>-.21**</td>
<td>-.09**</td>
<td>-.17**</td>
<td>-.15**</td>
<td>-.13**</td>
<td>-.18**</td>
<td>-.15**</td>
<td>-.01</td>
<td>.08**</td>
<td>-.03</td>
<td>.14**</td>
</tr>
<tr>
<td>27. AccET_T2</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>-.01</td>
<td>.06*</td>
<td>.03</td>
<td>.11**</td>
<td>-.03</td>
<td>-.04</td>
<td>-.07**</td>
<td>-.04</td>
</tr>
<tr>
<td>28. Guilt</td>
<td>-.28**</td>
<td>-.05</td>
<td>-.28**</td>
<td>-.28**</td>
<td>-.31**</td>
<td>-.31**</td>
<td>-.39**</td>
<td>-.05</td>
<td>-.09**</td>
<td>-.09**</td>
<td>-.10**</td>
</tr>
<tr>
<td>29. Fear</td>
<td>-.29**</td>
<td>-.06*</td>
<td>-.32**</td>
<td>-.30**</td>
<td>-.34**</td>
<td>-.35**</td>
<td>-.43**</td>
<td>-.06</td>
<td>-.09**</td>
<td>-.10**</td>
<td>-.10**</td>
</tr>
<tr>
<td>30. Anger</td>
<td>-.27**</td>
<td>-.05</td>
<td>-.25**</td>
<td>-.20**</td>
<td>-.30**</td>
<td>-.24**</td>
<td>-.31**</td>
<td>-.06</td>
<td>-.02</td>
<td>-.12**</td>
<td>-.01</td>
</tr>
<tr>
<td>31. Test Fairness</td>
<td>.27**</td>
<td>.11**</td>
<td>.32**</td>
<td>.25**</td>
<td>.31**</td>
<td>.26**</td>
<td>.30**</td>
<td>.11**</td>
<td>.07**</td>
<td>.14**</td>
<td>.04</td>
</tr>
</tbody>
</table>
Table 2 (Cont’d)

<table>
<thead>
<tr>
<th></th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Warning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>IM_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>BS_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>IM_FakerT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>BS_FakerT1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Faker_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>C_T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>IM_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>BS_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>C_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>E_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>A_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>O_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>ET_T2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>C_T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>E_T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>A_T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>O_T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>ET_T3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.84)</td>
</tr>
<tr>
<td>22.</td>
<td>AccC_T1</td>
<td>.12**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>AccC_T2</td>
<td>.02</td>
<td>.47**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>AccE_T2</td>
<td>.02</td>
<td>.22**</td>
<td>.31**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>AccA_T2</td>
<td>.03</td>
<td>.14**</td>
<td>.34**</td>
<td>.20**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>AccO_T2</td>
<td>.03</td>
<td>.16**</td>
<td>.31**</td>
<td>.43**</td>
<td>.18**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>AccET_T2</td>
<td>-.09**</td>
<td>.14**</td>
<td>.28**</td>
<td>.27**</td>
<td>.31**</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>Guilt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.94)</td>
</tr>
<tr>
<td>29.</td>
<td>Fear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.89)</td>
</tr>
<tr>
<td>30.</td>
<td>Anger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.91)</td>
</tr>
<tr>
<td>31.</td>
<td>Test Fairness</td>
<td>.12**</td>
<td>-.09**</td>
<td>-.10**</td>
<td>-.13**</td>
<td>.02</td>
<td>-.08**</td>
<td>.01</td>
<td>-.11**</td>
<td>-.16**</td>
<td>-.27**</td>
</tr>
</tbody>
</table>
Table 3.

Results of Testing Indirect Effects in Faker and Non-faker Groups Based on Bootstrapping

<table>
<thead>
<tr>
<th>Indirect effects</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fakers</strong></td>
<td><strong>Non-fakers</strong></td>
</tr>
<tr>
<td>Treatment → Guilt → Change in Score Accuracy</td>
<td>[.01, .13]</td>
</tr>
<tr>
<td>Treatment → Fear → Change in Score Accuracy</td>
<td>[-.04, .04]</td>
</tr>
<tr>
<td>Treatment → Anger → Change in Score Accuracy</td>
<td>[-.05, .02]</td>
</tr>
<tr>
<td>Treatment → Guilt → Perceive Test Fairness</td>
<td>[-.12, .11]</td>
</tr>
<tr>
<td>Treatment → Fear → Perceive Test Fairness</td>
<td>[-.11, .06]</td>
</tr>
<tr>
<td>Treatment → Anger → Perceive Test Fairness</td>
<td><strong>[-.19, -.03]</strong></td>
</tr>
</tbody>
</table>

*Note.* Bolded confidence intervals do not contain zero.
Figure 1. The Conceptual model.
Figure 2. The research design flow chart of the present study. BS = bogus statement; IM = impression management; WBI = the Work Behavior Inventory.
Figure 3. The multi-group path analysis results. Paths with two coefficients mean they are statistically different between fakers and non-fakers, with coefficients before and after slash for non-fakers and fakers, respectively. For the ease of presentation, the paths associated with control variables were not included in the figure. In addition, residuals associated with the three emotions were allowed to be correlated. * $p < .05$. ** $p < .01$. 
Appendix A

Messages Used in the Current Study

1. The mid-test warnings message

   Dear Candidate,

   Thank you for applying for graduate programs at Henan University and participating in this portion of the selection process. However, we have noticed some unusual response patterns in your answers and wish to clarify the issue. The personality inventory and the school activity survey which you are completing have two embedded social desirability scales. These scales identify people who might have tailored their responses to what Henan University wants to hear instead of present their true selves, in order to increase the chances of getting accepted.

   Your response profile up to this point is similar to that of someone who is known to be answering in a socially desirable way. We do not intend to insult your integrity; we only want to get a clear understanding of who you are. Inaccurate information from the personality test may lead to inaccurate and unfair assessment of your personality profile. Thus, we would like to underscore the importance of total honesty in completing these inventories.

   That said, we would like to offer you an opportunity to complete the inventories all over again. Remember, be yourself and answer each question as it best describes you. Finally, rest assured that your previous responses on these inventories will NOT be considered in our final selection decisions. However, we have found in the past that some candidates had repeatedly distorted their responses. These candidates were quickly discovered and were given zero grade in their personality test.

2. The control message

   Dear Candidate,

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

Six-month Test-retest Descriptives Based on an Independent New Graduate Student Sample

<table>
<thead>
<tr>
<th>Chinese WBI Dimensions</th>
<th>Test-retest $r$</th>
<th>Test Mean (SD)</th>
<th>Retest Mean (SD)</th>
<th>Change (Retest – Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extroversion</td>
<td>.80</td>
<td>2.99 (.52)</td>
<td>3.07 (.51)</td>
<td>.08</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.61</td>
<td>3.35 (.49)</td>
<td>3.54 (.36)</td>
<td>.19</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>.72</td>
<td>3.24 (.51)</td>
<td>3.33 (.49)</td>
<td>.09</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.73</td>
<td>3.36 (.38)</td>
<td>3.17 (.38)</td>
<td>-.19</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>.74</td>
<td>3.10 (.50)</td>
<td>3.16 (.47)</td>
<td>.06</td>
</tr>
</tbody>
</table>

*Note. n = 181. Both test administrations were for research purposes.*