A lab study comparing several faking measures under a simulated selection context

by

Yuan Yuan

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
December 12, 2015

Keywords: Faking, Impression Management, Self-deceptive Enhancement,
Bogus Statement, Over-claiming Questionnaire

Approved by

Jinyan Fan, Chair, Department of Psychology, Associate Professor of Psychology
Daniel Svyantek, Department of Psychology. Professor of Psychology
Alejandro Lazarte, Professor of Psychology

Abstract

Industrial and organizational psychologists have long been aware that faking is a concern in the personnel selection practices. In high-stakes selection context, applicants can be highly motivated to fake good on personality tests. However, researchers have not found a valid strategy to detect or capture applicant faking, and the validities of existing faking measurements have not been thoroughly examined. Meanwhile, researchers are arguing whether faking occurs in the selection context, and whether faking impacts the criterion-related validities of personality measures. The answers to the above two questions have been inconclusive, which is mostly due to the use of inappropriate research design and ineffective faking measures.

The present study is aimed to consolidate this literature through a systematic examination of the validities of several faking measures. Through a mixed design (i.e., both within-subjects and between-subjects design) with college students, the present study examines and compares the validity of four faking measures (i.e., the impression management inventory, self-deceptive enhancement inventory, bogus statement inventory and over-claiming questionnaire) under a simulated job application situation. The present study investigates and compares the validities of these faking measures through their correlations with the faking criterion as well as their moderation and suppression effects on the criterion-related validity of personality scores.

Table of Contents

List of Tables

List of Figures

Introduction

Personnel selection professionals have long been aware that individuals are highly motivated to fake good on non-cognitive measures under selection contexts. In high stakes personnel selection, faking means a tendency to distort responses to personality test in a socially favorably style. Rather than reporting their own personality accurately, fakers present the characteristics of 'ideal applicant' that they believe is socially approved. This responding style is referred to as Social Desirable Responding (SDR) (Zerbe & Paulhus, 1987), which describes individuals' tendency to represent themselves in a more favorably light than is accurate.

How to measure faking? On one hand, it is unreliable to depend on individuals' self-reported faking behaviors. On the other hand, the faking behaviors are too subtle to be observed or monitored by the third party, especially during the selection process. Faking researchers have agreed that it is better to define faking as individual score changes on personality tests between high motivated test situation (e.g., a selection context) and a low motivated test situation (e.g., a research context) (Fan et al., 2014; Griffith & Peterson, 2008). This definition of faking, which is also referred to hereafter as the "faking criterion", was originally adopted from the work of Paulhus and John (1998). The definition has helped researchers to observe faking accurately and more objectively.

It has been observed that at least some applicants would engage in faking under selection context, inflating test scores and impacting hiring decisions (Ross, Stecher, Miller & Levin, 1998; Griffith, Chmielowski &Yoshita, 2007; Griffith & Robie, 2012). In the study of Donovan, Dwight, and Hurtz (2003), it was reported that 32% applicants admitted exaggerating their positive attributes during the selection process, and 62% applicants agreed that they de-emphasized negative traits. Although researches have provided some evidence for the

occurrence of faking, IO psychologists are more concerned about whether faking negatively impacts the criterion-related validity of the selection tools (i.e., Ziegler, MacCann, & Roberts, 2012). The results are mixed in this regard. On one hand, with few exceptions, numerous studies have found that controlling for social desirability scores (i.e., from a scale assessing social desirable responding, or the SDR) does not improve criterion-related validities of personality scores (for meta-analyses, see Li & Bagger, 2006; Ones, Viswesvaran, & Reiss, 1996). On the other hand, a few researchers are not satisfied with the negative results of the suppression effect, and argue that faking may impact the validities of personality measures through the moderation effects (i.e., Berry, Page, & Sackett, 2007; Lanyon, Goodstein, & Wershba, 2014; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; White, Young, & Rumsey, 2001).

Explanation of the inconsistencies of faking literature has come in two ways. For one side, some contend that the mixed results are due to the fact that faking is not related to the test validities of personality measures (e.g., Ones et al., 1996). For the other side, some argue that faking is indeed detrimental to the selection system, and the inconsistent—especially negative results are due to the utilization of invalid faking measures and inappropriate research designs (Griffith et al., 2007; Griffith, Malm, English, Yoshita, & Guijar, 2006). To me, after in-depth literature review, I have come to the conclusion that the availability of valid faking measures is a prerequisite of studying whether faking negatively impacts the validities of personality tests (Griffith &Peterson, 2008). In order to answer the questions whether faking matters, researchers must have valid faking measures.

Although researchers have developed several measures for faking, their validities have not been fully examined. Traditionally, faking has been measured with SDR scales. The SDR

scales adopt an "unlikely virtues" approach and are composed of items describing highly desirable, yet unlikely virtues. A typical SDR item is "I never throw litter on street", or "I never regrets on my decisions". Individuals who endorse too many socially desirable items are considered as potential fakers.

Up to the date, many SDR scales have been developed, including the Edwards Social Desirability scale (Edwards, 1957), the Marlowe-Crowne scale (Crowne & Marlowe, 1960), the Good Impression scale of CPI (Gough, 1952, 1987; Gough & Bradley, 1996), the MMPI Lie scale (Meehl & Hathaway, 1946), as well as the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1988, 1998, 2006). Under high-motivated context, the scores on SDR scales would increase substantially. Research showed that under faking context, the amount of score change on SDR scales is greater than that of personality measures (Viswesvaran & Ones, 1999). However, SDR scales seem unable to impact the criterion-related validities of personality measures (Ones et al., 1996; Li & Bagger, 2010). Despite of the popularity of SDR scales, these findings cast doubts on whether the SDR scales are valid faking measures as they were assumed.

More recently, researchers developed a 'false-item' approach to measure faking. In this method, faking is tested in a survey containing both false items and genuine items. Those who feign knowledge or experience of nonexistent items are identified as fakers. Examples of the "false-item" approach include the bogus statement inventory (BS inventory; Anderson, Warner, & Spencer, 1984) and the over-claiming questionnaire (OCQ; Paulhus, Harm, Bruce & Lysy, 2001). Under high-motivated test condition, the endorsement of false items on BS inventory or OCQ would increase substantially (e.g., Fan et al., 2012; Paulhus et al., 2003). However, in general, there have not been many faking studies using BS inventories or OCQ. Although the BS inventory and the OCQ have showed some promises in impacting the criterion-related

validity of personality measures (e.g., Anderson et al., 1984; Bing Kluemper, Davison, Taylor & Novicevic, 2011), their validities are far from being fully established.

The present study is aimed to consolidate the literature through a re-examination and comparison of the validities of faking measures. Under a simulated job application situation, the study examines and compares the validity of four popular faking measures: the impression management scale (IM; Paulhus, 1988, 1998), self-deceptive enhancement scale (SDE; Paulhus, 1988, 1998), the bogus statement inventory (BS, Dwight & Donavan, 2003) and the foil index of over-claiming questionnaire (OCQ-Foils; Bing et al, 2011). Specifically, the study examines whether these four faking measures (a) are responsive to high-motivated test context (i.e., how a faking measure would change under the impacts of faking in regards to its features such as mean scores, correlational patterns, or reliabilities); (b) correlate with residual personality scores in a within-subjects design; (c) suppress personality validities; and (d) moderate personality validities. In the following sections, I will introduce the history of faking research in organizational settings, and then review the four faking measures that I will use (i.e., IM, SDE, BS and OCQ-Foils), discussing issues regarding their development, evidences, and validities. Finally, the hypotheses to examine and compare the validity of four faking measures will be presented, and methods and procedures will be described.

History of Faking Research in Organizational Settings

Faking has long been a concern in personnel selection field. Personality inventories become more and more popular as a selection tool. Researchers have been aware that the self-report personality items are transparent, and thus prone to fake (Griffith & Peterson, 2008). Faking research under organizational settings has been dealt with two fundamental questions, i.e., whether faking occurs, and whether it matters. For both topics, field studies yield mixed

findings. This section tries to clarify the two questions through a re-examination of the faking literature.

Whether faking occurs? Motivated individuals are capable of manipulating their scores on personality tests. Viswesvaran and Ones (1996) showed that in lab settings, participants who received the "fake-good" instructions scored substantially higher on personality measures than participants under a control condition. However, do job applicants engage in faking in the real life selection context? This question has been investigated with both between-subjects studies and within-subjects studies.

Most between-subjects studies found positive results regarding the occurrence of faking. In field studies, researchers observed that job applicants scored significantly higher on personality tests than students or job incumbents (Barrick & Mount, 1996; Elliot, 1981; Schmit and Ryan, 1993; Stokes, Hogan, & Snell, 1993). Rosse et al (1998) reported that job applicants scored one-standard-deviation higher in personality tests than job incumbents. In a more recent meta-analysis study investigating faking among job applicants, Birkeland, Manson, Kisamore, Brannick & Smith (2006) revealed that job applicants scored significantly higher on all Big-Five dimensions of personalities than non-job-applicants, with the effect sizes ranging from .11 to .45. These results clearly indicate that in the real-life selection context, faking does occur, and at least some applicants engage in faking under selection context.

Although between-subjects studies generated positive results on the occurrence of faking, research has revealed that within-subjects studies tend to produce more accurate estimates than between-subjects designs (Viswesvaran & Ones, 1996). In within-subjects studies, the results regarding the occurrence of faking are mixed. On one hand, majority of the within-subjects studies supported the occurrence of faking. Several test-retest studies using job applicants

showed that applicants scored much higher on personality tests when they were tested under real-life selection context, comparing to under a research context.

Landers, Sackett and Tuzinski (2011) found that internal applicants for managerial positions who had failed a personality test the first time increased their personality scores substantially at the retest. Griffith et al. (2007) found that when applicants volunteered for a retest a month following their job application, they scored much lower on the same scales than when they were applying for the job. Using a confidence interval (CI) method (for more details about the CI, see Dudek, 1979), Griffith et al. (2007) found that among those who volunteered in the re-test, 30-50% of them were engaging in faking under the application situation. Also in a test-retest design using a similar CI method, Arthur, Glaze, Villado, and Taylor (2010) concluded that roughly one third of applicants faked on personality tests when they took it for the first time under the selection context.

On the other hand, some within-subjects studies found opposite results towards the occurrence of faking. For instance, Hogan, Barrett, and Hogan (2007) reported that job applicants who were denied employment due to their failure of passing a test battery had little score elevation on the personality test when they were given the chance to retake the same test battery 6 months later. The authors argued that although these job applicants supposedly should have been much more motivated to fake at the retest, faking did not actually occur. In another study, Ellingson, Sackett, and Connelly (2007) identified a large group of professionals who took the California Psychological Inventory (CPI) twice, yielding four groups (selection/selection, selection/development, development/selection, and development/development). Ellingson et al. concluded that the degree of faking from the development context to the selection context was minimal across all personality dimensions.

Although the above studies did not support the occurrence of faking, Fan, Zhao, Li, Wang, and Meng (2014) pointed out that other factors can be responsible for the negative results. In Hogan et al. (2007)'s research, Fan et al. (2014) noted that because several tests were administrated to the applicants all-together, the applicants who took the retest may choose to improve the scores on other measures (e.g., a cognitive measure), rather than on the personality measure. Fan et al. (2014) also pointed out that for the Ellingson et al. (2007) study, Ellingson et al. reported that applicants faked on a few personality dimensions. However, Ellingson et al. took an average approach on the amount of faking across all personality dimensions, which mitigated the overall effects of faking (Fan et al., 2014).

In summary, the literature review shows that although there are some inconsistencies regarding the occurrence of faking, plenty evidences have indicated that faking does occur. When excluding the potential confounders brought by research design and data analysis skills, it is clear that most between-subjects designs and within-subjects designs indicated that faking does occur. Moreover, a substantial proportion of applicants engaged in faking under the selection context. Given these evidences, it is meaningful to discuss the following questions, i.e., whether faking matters, and how to measure faking.

Whether faking matters? Given the occurrence of faking, some researchers are worried that occurrence of faking can negatively impact the criterion-related validity of non-cognitive measures. Yet a conventional wisdom in the faking has been that, even if faking occurs, it does not matter. This view is based on the fact that numerous studies (including meta-analyses) have shown that controlling for SDR scales did not improve criterion-related validities of personality scores, indicating a lack of suppression effects (e.g., McGrath et al. , 2010; Li & Bagger, 2006; Ones et al., 1996). However, a closer examination of the literature discovered that the

conventional view is biased in several aspects. As a result, a re-examination of the topic "whether faking matters" is necessary.

To begin with, the conventional view was based on the fact that most studies did not find a suppression effect of faking measures. However, many of these studies are flawed in their way of testing the suppression effects. Berry et al. (2007) and McGrath et al. (2010) pointed out that most previous studies did not test suppression effects in a statistically appropriate manner. In the past, many researchers used the semipartial correlation to test the suppression effect, which is inappropriate. In fact, the most appropriate statistic test for suppression effect is regression analysis, as indicated by Bing et al. (2011) in their study.

Another problem associated with previous faking research is the research design. Given the difficulty to find real job applicants, organizational researchers tend to conduct faking studies with college students or job-incumbents. As a result, many faking studies were conducted under a condition in which participants lack a real motivation to fake (e.g., Ganster, Hennessey and Luthans, 1983; Borkenau & Ostendorf, 1992). These researches typically found negative results towards the suppression effect of SDR scales. However, one should not expect to see the suppression effects of faking measures under a context in which faking does not exist. As a result, the conclusion that faking does not suppress the prediction of personality tests generated from these studies can be biased and misleading.

In fact, a few studies that have overcome the weaknesses mentioned above actually found a suppression effect of faking measures. In the study of Bing et al. (2011), it was found that under an instructed faking condition, the foils index of OCQ could suppress the prediction of achievement striving (a sub-dimension of school-specific conscientiousness, see Schmit, Ryan, Stierwalt & Powell, 1995) on college students' GPA. Anderson et al. (1984) also reported

significant suppression effects of a BS inventory upon the prediction of a self-reported job experience assessment among a sample of job applicants. This evidence further indicates that when tested properly, the suppression effects of faking measures can be found.

Furthermore, the conventional view inappropriately equates the lack of suppression effects with the statement that "faking does not negatively impact test validities". In fact, faking's impact on validity can also be indicated by the moderation effects. The moderation effects indicate a difference in the criterion-related validity of personality measures across samples (e.g., job-applicants samples vs. incumbent's samples) or conditions (e.g., application context vs. research context). As Berry et al. (2007) pointed out, most previous researchers failed to consider the possibility that faking measures may moderate. Several meta-analysis studies also overlooked the moderation effect (Ones et al., 1996; Li & Baggers, 2006). Not until recently, the moderation effects of faking measures are raised up and supported by a few studies (i.e., Berry et al., 2007; Hough et al., 1990; White et al., 2001; Lanyon et al., 2014). Although the moderation effects have gained some initial supports, the studies on the moderation effect have been rare. To better support the moderation effects, more studies are needed in this direction.

In addition, the conventional view overlooked the evidence that faking can alter hiring decisions via impacting ranking orders. Fakers tend to dominate the top ranking in high stake personnel selection context, which will result in the hiring of potential fakers, rather than the honest individuals (Peterson, Griffith & Converse, 2009; Fan et al., 2012, Griffith et al., 2007). These results warn us that faking impacts the utility of the selection system, which makes it necessary to detect and control for faking.

Finally, the debate over faking's impact on test validities highlights the importance of re-examining the validities of faking measures. Previous researchers have based their results on the assumption that the faking measures they used can capture faking effectively. However, this assumption may not be true. For a long time, there have been much arguments regarding whether SDR scales are appropriate measures of faking (e.g., Smith & Ellingson, 2002; Holden & Passey, 2010; Uziel, 2010). On the other side, there has not been much literature behind the validities of BS inventories and OCQ. The validity of faking measures plays an important role in assessing the impact of faking on the validity of selection systems, a topic to which I will turn in the next section.

In summary, the literature is still not ready to answer the question "whether faking matters", because previous researchers did not design studies carefully, did not observe faking under high-motivated context, and did not analyze data correctly. Here I summarize a few points that researchers should follow when conducting faking studies: 1) test the effects of faking upon validities using appropriate methods, and 2) under high motivation contexts; 3) incorporate moderation effects of faking into research plan; and 4) use valid faking measures. The faking literature should have been less inconsistent if majority of faking studies had followed the above points; nevertheless, given the inconclusive of the current faking studies, it is time to re-evaluate the old questions in more advanced methods, as well as with more effective faking measures.

Validation of Faking Measures

Up to the date, hundreds of studies have made the attempt to study faking using one or more faking measures (Ones et al., 1996; Li & Bagger., 2006; McGrath et al. , 2010). However, few studies have ever examined whether the faking measures that they used are indeed valid. Although there are several methods to validate faking measures, these methods have not been summarized and compared against each other. This section will review and explain several ways to examine the validities of faking measures. Those validation methods are used in the present study as well.

First of all, a pre-requisite of valid faking measures is to be sensitive, or responsive, to the high-motivated test context in which faking occurs. As mentioned earlier, responses to test context refers to the ways that a faking measure would change in regards to its features (i.e., mean score, correlational pattern with other constructs, or reliabilities) under a high-motivated test context. The most important reaction of faking measures to the test context lies in the change of mean scores. Under high-motivated test contests, the means of all faking measures would increase. For SDR scales, the endorsement of unlikely virtue items increased under faking condition (see Viswesvaran & Ones, 1999). For BS inventories or OCQ, the endorsement of bogus items or foils also increased when the test motivation is high (for BS inventory, see Dwight & Donovan, 2003; Fan et al., 2012; for OCQ, see Paulhus et al., 2003; Bing et al., 2011). Moreover, faking measures can also react to test context through changes in correlational patterns with other constructs. For example, under high-motivated test condition, IM and SDE tend to show higher correlation with each other, as well as with personalities (Paulhus, 2002). Finally, some faking measures also react to the test context through changes in reliabilities (e.g., the BS inventory; see Fan et al., 2012). The specific reactions of each faking measure to the

faking context will be discussed in more details in the next section, when the faking measures are reviewed.

Another way to verify the validity of faking measures is through their correlations with the faking criterion—the observed within-subjects score change across high-versus-low motivated contexts (Griffith & Peterson, 2008). A significant and substantial correlation between a faking measure and the score change would prove the faking measure to be valid, while a non-significant correlation renders the faking measure invalid. The within-subjects score change is adopted from the work of Paulhus and John (1998), who defined SDR through the residual score of self-rated Big-five personalities over peer-rated Big-Five personalities. The residual between self-ratings and other ratings is also referred to as 'self-other' residuals. Although similar, the within-subjects score change are more close to applicant settings, and is more suitable for faking research in IO psychology.

The validity of faking measures can also be indicated by their impacts upon the criterion-related validity of personality measures. Researches have shown that the criterion-related validities of personalities for job applicants are lower than job incumbents or students sample, which indicates a moderation effect of contexts (Mueller-Hanson, Heggestad, & Thornton,2003; White et al., 2001). Moreover, Peterson, Griffith, Isaacson, O'Connell, and Mangos (2011) also reported a significant moderation effect of the status of fakers. In Peterson et al. study, a sample of real applicants completed a selection battery online for manufacturing positions. After the application procedure, the applicants were invited to participate in a follow-up study that was independent from the selection procedure and unrelated to the hiring decisions. A small proportion (i.e., 6%) of the job applicants participated in the follow up study, which occurred six weeks later. In the follow up study, participants filled out the same conscientiousness scale that

had been included in the selection battery and self-reported their counterproductive workplace behaviors (i.e., CWBs). The faking was assessed with the Marlowe-Crowne (Crowne & Marlowe, 1960) scale and also the faking criterion. The faker's status was determined by the amount of within-subjects score change on conscientiousness between the selection context and the follow-up study. The results showed that the relationship between conscientiousness and the CWB for fakers identified via the faking criterion was substantially lower than that of non-fakers, which indicating a significant moderation effect. The Marlowe-Crowne scale, however, failed to show any impact on the criterion-related validity of personality tests.

Researchers have also found that controlling for faking measures can improve the criterion-related validities of selection tools substantially, which indicates a suppression effect (Anderson et al., 1984; Bing et al., 2011). With a sample of job, Anderson et al. (1984) found a significant suppression effect of an inflation scale (i.e., BS inventory) upon the prediction of a self-reported job experience assessment towards an objective job performance test (i.e., a typing test). Under a simulated application situation, Bing et al. (2011) also found a suppression effect of OCQ-Foils upon the prediction of conscientiousness towards students' cumulative GPA. Based on the evidences, one can make the inference that if a measure captures faking effectively, it should demonstrate either moderation effects or suppression effects on the criterion-related validity of personality measures.

Since the present study is aimed to examine and compare the validity four faking measures (i.e., IM, SDE, BS, and OCQ-Foils), the next section will introduce these faking measures and review their validities based on the validation methods mentioned above. For each faking measure, its validity will be discussed on its reactions to faking contexts, whether it

correlated with the faking criterion and whether it moderates or suppresses the criterion-related validity of personality measures.

Review of Faking Measures

      Impression management (IM) and self-deceptive enhancement (SDE). The IM and SDE are two subscales of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1988, 1998, 2006). As typical SDR scales, IM and SDE capture faking through the endorsement of unlikely virtue items. A typical IM item is "I always tell the truth", while a typical SDE item is "I never regret on my decisions". Up to the date, the IM and SDE are the most popular measures for SDR. The popularity of IM and SDE has led to their wide application in the faking research at workplaces and under the selection context.

      Development. Researchers revealed that the concept of SDR is not a unitary dimension (Wiggins, 1964; Damarin & Messick, 1965). In 1984, Paulhus conducted factor analysis over several existing SDR measures and reconfirmed the two-factor structure: as previously reported, there were an alpha factor, and a gamma factor. According to Paulhus (1984), the MMPI Lie scale (Meehl & Hathaway, 1946), Wiggins's social desirability scale (Wiggins, 1959), and Gur's other-deception scale (Sackeim & Gur, 1978) loaded onto the same factor, while the Edwards's social desirability scale (Edwards, 1957), Byrne's repression-sensitization scale(Byrne, 1961), and Gur's self-deception scale(Sackeim & Gur, 1978) were loaded onto the other factor. Paulhus renamed the two factors as impression management (IM) and self-deceptive enhancement (SDE) base on his interpretation of the factors. Based on his findings, Paulhus developed a SDR scale named the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1988, 1998), which include separate measures for the two dimensions—a subscale for IM, and another for SDE.

Interpretation. Originally, Paulhus (1984) defined IM as an intentional distortion of self-descriptions to gain approval from others. SDE, on the other hand, was defined as an unconscious propensity to view oneself in a favorable light. This interpretation differentiates IM from SDE on the level of consciousness. However, this original definition has been challenged by empirical studies. Researches have shown that under factor analysis, the BIDR tend to split into three factors, rather than two (see Paulhus, 2002). Researchers have found that under motivated context, the scores of both IM and SDE increased significantly than honest context, which indicating that SDE can also be "intentional", rather than non-intentional (see Paulhus, 2002).   In order to integrate those findings, Paulhus later changed his view on the nature of IM and SDE. The IM is redefined as communal bias, for which individuals claim "saint-like" traits and exaggerate virtues. The SDE is redefined as agentic bias, in which individuals claim "super-heroes" traits and exaggerate intelligence and abilities (see Paulhus, 2002). In other words, Paulhus believes that the difference between the IM and SDE actually lies in the content (i.e., communal vs. agentic), rather than the level of consciousness (i.e., intentional vs. unintentional).

Reactions to high-motivated test contexts. Typically, the scores on both IM and SDE tend to increase substantially under high-motivated test conditions. This phenomenon has been observed in lab studies under faking instructions (Pauls & Crost, 2004; Holden, 2008) and under simulated application situations (Dodaj, 2012; Holden & Evory, 2005; Bing et al., 2011). Moreover, individuals can also inflate their scores on IM or SDE differentially based on explicated instructions or environmental demands (Paulhus & Notareschi, 1993; Pauls & Crost, 2004; Dodaj, 2012). When applying for a simulated job position, participants scored higher on IM than SDE when applying for a teacher's position, while scored higher on SDE than IM when applying for a manager's position (Dodaj, 2012). These evidences indicate that in general, both

IM and SDE are sensitive to test motivation. Moreover, there are also job contents that

differentially sensitive to IM and SDE. All in all, it is possible that IM and SDE can both

capture faking and capture faking differentially under high motivated context.

Under high motivated test condition, IM and SDE also tend to produce higher correlation

with each other. According to Paulhus (2002), the correlation between IM and SDE is

around .30 under the honest condition, which would increase to .50 or higher under high

motivated context. Moreover, high-motivated context also impacts the reliabilities of IM and

SDE. IM and SDE typically show quite high reliabilities under low-motivated context (i.e., .51

for IM, .63 for SDE; Bing et al., 2011). Under high-motivated condition, the reliabilities of IM

and SDE tend to increase slightly than in low-motivated condition (i.e., .71 for IM, .73 for SDE;

Bing et al., 2011). The reliability change of IM and SDE under different test context has been

well researched previously, and it is not the main focus of the present study.

Correlation with the faking criterion. The only study that has examined the relationship

between IM/SDE and the faking criterion is from Griffith et al. (2006). In a within-subjects

study, Griffith et al. administered a personality test twice with a sample of college students under

high-and-low motivated test context. The high-motivated test context was created using a

deception technique. In the study, although the personality tests were administered twice under

both high-and-low motivated contexts, the BIDR scale (i.e., IM and SDE) was only administered

to participants only after the debriefing, which is a low motivated context. Griffith et al. reported

no significant correlation between IM/SDE and the faking criterion. As mentioned earlier, faking

did not occur under an honest condition. Therefore, the results from Griffith et al. (2006) are not

convincing that the IM and SDE are not valid faking measures.

In another study, Peterson et al. (2011) investigated the correlation between the Marlowe-Crowne scale and the faking criterion. The Marlowe-Crowne scale administrated under the application context showed no significant correlation with the faking criterion. Although Peterson et al. found a negative result towards the proxy validity of the Marlowe-Crowne scale, this result has not been replicated with IM/SDE under high motivated test condition yet. More research is needed to investigate IM/SDE correlations with the faking criterion appropriately.

Suppression effects. Meta analyses studies concluded negative results regarding the suppression effects of SDR scales. Li and Bagger (2006) summarized 236 articles exploring the relationship between IM/SDE and Big-Five personalities and job performance. Li and Bagger found that in general, controlling IM/SDE does not improve the correlations between personalities and job performance. McGrath et al. (2010) used a section to focus on the use of SDR scales at workplaces and their impact on the criterion-related validity of personalities. Among the 11 articles incorporated, 10 used some SDR scale to measure faking, and none of them found significant suppression effects. Ones et al. (1996) also concluded that partialling out SDR does not increase the criterion-related validity of personalities. Although previous studies failed to observe significant suppression effects of IM/SDE upon the prediction of personalities, it is important to interpret the results critically and consider the fact that most previous studies were conducted under low motivation context, in which faking did not occur.

Moderation effects. Comparing with the suppression effects, fewer efforts were made in the research of potential moderation effects of SDR scales. The meta-analyses of Li and Bagger (2006) and Ones et al. (1996) did not look at the moderation effects. In a more recent meta-analysis study, McGrath et al. (2010) found seven studies that tested the moderation effects of

faking measures upon the prediction of personalities. Among them, only Hough et al. (1990) found significant moderation effects.

Indeed, I was only able to locate four published studies that found significant moderation effects of SDR scales (i.e., Hough et al., 1990; White et al., 2001; Berry et al., 2007; Lanyon et al., 2014). Hough et al (1990) found a significant, yet very weak moderation effect of SDR scores. Using a concurrent design with a large military sample, the personality validities for soldiers with a high SDR score were only slightly smaller than personality validities for soldiers with a low SDR score (Hough et al., 1990). Although the moderation effects of Hough et al. study were relatively weak, the other three articles found stronger results. Using a prospective design, White et al. (2011) found a significant moderation effect of SDR scales on personality validities in another military sample. In a concurrent design involving a managerial sample, Berry et al. (2007) found that SDE moderated the Extroversion – performance relationship such that personality validity was minimal when SDE was high, but was positive when SDE was low. However, Berry et al. did not find IM's moderation effects. The Lanyon et al. (2014) article found moderation effects of the Good Impression scores in two out of four incumbent samples (i.e., nurses and retail store managers) in the expected direction such that the criterion-related validity is higher when SDR is low, and is lower when SDR is high.

In general, the above four articles provided some support of the moderation effects of SDR scales. However, only Berry et al. used IM/SDE as indicators of faking; the rest of studies used other SDR scales. Although there are some limitations in the above studies (i.e., not using real job applicants sample; used concurrent design; tested the moderation effects inappropriately, etc.), a more in-depth investigation of the moderation effects is worthwhile, and more studies are needed to examine the moderation effects of IM/SDE scores.

Issue and comments—substance vs. style. Researchers have reported a moderate

correlation between IM/SDE and personality measures under low motivated context (e.g., Bing

et al., 2012; Holden & Passey, 2010; Pauls & Stemmler, 2003), which may even increase under

high-motivated context (i.e., Smith & Ellingson, 2002; Bing et al, 2011). Since the IM/SDE

scales correlate moderately with personality traits, some researchers hold the opinion that

IM/SDE may actually measure meaningful traits, rather than faking (Uziel, 2010, 2014). Yet

there are some inconsistencies regarding what type of traits the IM/SDE scales reflect. In

regarding to the IM scale, Uziel (2010) reviewed several articles in the social-personality area

and concluded that the IM actually reflects prosocial traits due to that it could predict

interpersonal social adjustment. In contrast, Davis, Thake and Weekes (2012) showed that

among a sample of male offenders, IM associates with more sever crime and longer periods of

sentences, which leads to the conclusion that the IM, rather than SDE, associates with anti-social

behaviors. Meanwhile, Paulhus neither denied nor confirmed the issue regarding "content vs.

style". Paulhus pointed out that although there are some evidences for both sides, the argument

is still not conclusive, and more researches are needed in this regard (Blasberga, Rogersa &

Paulhus, 2013).

In regarding to the debates over "substance versus style", I also have some comments. In

a way, it seems that the SDR scales focus on the "good", as well as "fake". Yet this is the only

possible way to catch applicants faking while they are "fake good". It reasonable that SDR

scales may reflect some good traits while measuring fake good, but for IO psychologist, the fact

that SDR reflects "faking" is more important than whether it measures "trait". To begin with, it

has been repeatedly shown by meta-analysis studies that the IM/SDE has no main effects upon

job performances (Li & Bagger, 2006; Ones et al., 1996). As a result, even under the assumption

that SDR scales do measure traits— the traits have been proved to be irrelevant to job performances. Furthermore, more and more evidences are showing that faking does matter in the selection field. In the study of Peterson et al. (2011), the amount of faking operationalized by score change across situations was directly and positively related to self-reported counter productive work behaviors (CWB), while the score on Marlow-Crown scale was not. The results indicated that faking could directly impact some type of job performance, which provide additional reasons why it is necessary to investigate faking.

All in all, the above evidences have indicated that for IO researchers, the question to pursue is how to measure faking more effectively, rather than whether faking measures also capture some traits. Since the present study is not concerned of whether IM/SDE scales measure personality traits, I will only focus the first question, i.e., whether IM/SDE captures faking given the motivation.

The bogus statement (BS) inventory. The idea of BS inventory was firstly used in researches with biodata measures. Since the 1960s, biodata researchers try to mix bogus items together with biodata items to control for potential faking on bio-data measures (Brumback, 1969; Goldstein, 1971; Klien and Owens, 1965; Owens, 1976). Although there used to be only a few bogus items added, those items are effective in identifying potential fakers and improve the criterion-related validity of bio-data measures (Anderson et al., 1984; Pannone, 1984; Levashina , 2009). The success of bogus items in biodata measures leads to the use of BS inventory in the personality tests (Dwight & Donnavan, 2003).

Structure. The BS inventories resemble task experience surveys and are composed of both genuine tasks and a few (i.e., one or two) bogus tasks. The genuine tasks are real and related to the job content. The bogus tasks are false and are superficially similar to the genuine

tasks. For example, when used for academic selection settings, a real task on the BS inventory could be "utilize scientific calculator", which is related to the academic criterion. Meanwhile, a bogus item could be "utilize murray-web system to locate unpublished research articles", in which the "murray-web" is not a website for unpublished articles. A rating scale is used for respondents to indicate their frequency of engaging with these tasks in the past. An over endorsement of the bogus tasks are viewed as signs for faking.

Reactions to high-motivated test contexts. The endorsement of bogus items is higher under high-motivated context. Fan et al. (2012) administrated a two-item bogus statement scale to a sample of college students in China. The students were randomly assigned to either a simulated job application condition or a control condition. Under honest condition, the endorsement rate of more than one BS items is zero. While under control condition, the endorsement rate of more than one BS items is 15%.

Moreover, bogus statement items demonstrate higher reliabilities under high motivated conditions (Anderson et al., 1984; Fan et al., 2012). Fan et al. (2012) reported that Cronbach's alpha of a two-item BS inventory is .60 under a simulated job application condition, which dropped to .07 under a control condition. Consistent with Fan et al.'s results, Anderson et al. (1984) also reported that the Cronbach's alpha for BS inventories of ranged from .76 to .91 among a sample of real job applicants. These evidences indicate that under high-motivated condition, the BS items have strong internal consistency.

The reason to conduct the reliability test came in two ways. To begin with, since the contents of bogus items are false and non-existent, there is a concern regarding whether the bogus items are capable of capturing any meaningful construct, or they are "merely tricky or confusing" (p. 578, Anderson et al., 1984). High internal consistency of BS inventory under

high-motivated context reveals that the bogus items do capture some construct (Anderson et al., 1984).   Moreover, if BS inventory shows high reliability under high-motivated context, one can then claim that the BS inventory is capturing faking or response biases, considering the extremely low internal consistency that BS inventory shows under low motivated context.

Researches provided further evidences regarding the construct validity of BS inventories under faking context. Evidences showed that the scores on BS inventory are moderately and positively related to other performance indicators (i.e., bio-data measures and personalities; Levashina , Morgeson, & Campion, 2009; Anderson, 1984; Fan et al., 2012). In the study of Levashina et al. (2009), applicants' scores on three BS items showed low, yet significant correlations with their scores on the biodata survey (i.e., r=.06, .15, .19, all *p*<.0001.). In Anderson et al. (1984) research, the job applicants' BS scores showed significant positive correlation with their scores on a self-reported task experience assessments. In Fan et al. (2012) study, the applicants who endorsed more than one bogus item on the BS inventory also obtained a higher mean on job-relevant dimensions of 16PF personality scales. These findings are consistent with the assumption that the BS inventory is an indicator of faking. This is due to the fact that fakers typically score higher on personality and bio-data measures than non-fakers.

Correlation with faking criterion. Previous studies have not examined the relationship between the BS inventory and the faking criterion. It is still unknown whether the BS scores tested under a high-motivated condition would correlate with the faking criterion.

Suppression effect. Anderson et al. (1984) found a significant suppression effect of the BS inventory among a job applicants sample. In the study, the BS inventory increased the prediction of a self-assessment exam towards a job performance criterion substantially and

explained an additional 16% of variances. This is the only study that reported the suppression effect of BS inventory. More researches are needed in the direction.

Moderation effect. Pannone (1984) found a moderation effect of bogus items among a sample of applicants for electrician positions. In a real life application situation, job applicants completed a written test as well as a biographical survey assessing their past performance in electrical task domain. The written test contained one bogus item that was aimed to detect fakers. As a result, 65.2% of test takers endorsed the bogus item, which indicating an exaggeration in past experiences. For those who did not endorse the bogus item, the correlation between the biographical measure and the written test was .55 (p<.01). This value dropped to .26 (p<.05) for those who endorsed the item, which indicating a moderation effect. This is the only study that reported a moderation effect of BS inventory. More researches are needed in the direction.

Issues and comments--relationship with cognitive ability. Under high-motivated context, scores on BS inventory tend to correlate negatively with abilities (i.e., work knowledge and verbal ability) and job performance (Levashina et al., 2009; Anderson et al., 1984). Under application situation, the scores on BS items were negatively related with job knowledge and verbal abilities, as well as a simulated job performance test (Levashina et al., 2009; Anderson et al., 1984). These results are congruent with the idea that individuals with less knowledge and skills are more likely to fake in order to make up for their deficiencies. This argument is also supported by the study of Paulhus and Dubois (2014), which showed that knowledge self-enhancing is detrimental to overall performance.

Aside from the above results, it is also revealed that among those who engaged in faking, individuals with high abilities faked more than individuals with low abilities (Levashina et al.,

2009). This is due to the fact that those who high in abilities are more capable of faking and thus

will obtain a higher score (Levashina et al., 2009).

In general, the studies on BS inventories have been rare, and even less studies have

examined the relationship between the BS inventories and cognitive abilities. Although it is still

premature to conclude a relationship between BS and cognitive abilities, this topic is important

since such a relationship can cause adverse impacts. Therefore, the current study will also

investigate whether the scores on the BS inventory is related with the cognitive ability and

overall academic performance (i.e., as indicated by students' ACT/SAT, cumulative GPA,

respectively).

The over-claiming questionnaire (OCQ). The OCQ was firstly introduced into faking

research by Paulhus and Bruce (1990). The OCQ includes both genuine items and nonexistent

items (i.e., foils) that are related to a specific subject (e.g., academia, music, sports). The earliest

version of OCQ—the OCQ-150—taps on 10 of academic subjects (Paulhus, Harms, Bruce, &

Lysy, 2001). A total of 150 items are included, with15 items for each specific subject (i.e.,

history, fine arts, or physical science). For each subject, 3 out of the 15 items are foils.

Participants are asked to rate their familiarity with both genuine items and foil items.

The OCQ is scored using the signal detection techniques (SDT; Swets, 1964). According

to SDT, OCQ can be scored into two indices (i.e., the accuracy, bias). The accuracy index

represents test takers' ability to distinguish between genuine items and foils, which is an

indicator of cognitive ability. The bias index represents the test-takers tendency to say 'yes' to

all questions, which is an indicator of knowledge exaggeration. The bias index has also been

used as an indicator of faking under high-motivated test conditions (Paulhus et al., 2003).

Construct. According to the SDT, the calculation of accuracy and bias indices is based on two proportions (i.e., hit rate, false alarm rate). The hit-rate refers to the ration between participants claimed genuine items and all genuine items. The false-alarm rate refers to the ratio between claimed foils and all foils. Specifically, the accuracy index is calculated by the difference between the hit-rate and the false-alarm rate, while the bias index is calculated by the average of the hit-rate and the false-alarm rate. The accuracy index and the bias index are negatively and moderately correlated, yet the correlation can vary under different test situations (Paulhus & Harms, 2004; Paulhus et al., 2003).

An additional way to score OCT surveys is through direct counting of endorsed foils (i.e., the foil index). First introduced in the study of Bing et al. (2011), the foil index has been recognized as another index of knowledge exaggeration (Paulhus, 2011).

Indices. The accuracy index is an indicator of knowledge and global cognitive ability (Paulhus, 2011). The accuracy index of academic OCQ showed moderate-to-high correlations with measures of knowledge (i.e., Wonderlic IQ test, the UBC Word test, students' cumulative GPA; Paulhus & Harms, 2004; Paulhu & Dubois, 2014). The accuracy index showed no correlation with any Big-Five personality trait (Zigers, Kemper & Rammstedt, 2013; Bing et al., 2011). These results supported the conclusion the accuracy index is a measure for cognitive ability, aside from personalities.

The bias index contains three elements (i.e., self-enhancement, cognitive bias, and motivation; Williams, Paulhus, & Nathanson, 2002; Nathanson & Paulhus, 2005; Paulhus et al., 2003). Under low-motivated context, the bias index of academic OCQ correlates positively with measures of trait self-enhancement such as SDE and Narcissism Personality Inventory (Raskin & Hall, 1988; Paulhus et al., 2003; Randall & Fernandes, 1991; Paulhus & Williams, 2002; Paulhus

et al., 2003; Tracy, Cheng, Robins, & Trzeszniewski, 2009). Moreover, the bias index also correlates with self-enhancing related personality traits. Zigler et al. (2013) found that the bias index was positively correlated with extraversion and openness while was negatively correlated with neuroticism. These results indicate a self-enhancement component in the bias index.

Doubts have been casted regarding whether the endorsement of foils was due to the knowledge exaggeration, or merely due to a memory bias. Research has revealed that those who are high on OCQ bias index also tend to have a global memory bias (Williams, Paulhus, & Nathanson, 2002). However, the correlation between bias and self-enhancement sustains even after controlling for the memory bias (Williams et al., 2002). These results indicate that although memory bias contributes to the endorsement of foil items, the self-enhancement element sustains above and beyond the memory biases.

The endorsing of foils is also impacted by individual motivation and situational cues. In a study about the OCQ for non-academic topics, the bias index only correlated with narcissism on the topics that the respondent values (Nathanson & Paulhus, 2005). The results indicate that people only fake on the things they care about, and motivation is a powerful agent of faking. Moreover, Paulhus et al. (2003) observed that under high-motivated context, the scores on the bias index increased substantially and gained correlations with IM. The fact that the bias index reacts to faking indicates that it also reflects situational cues.

The foil index is a straight forward way to score faking (Bing et al., 2011). The foil index has been adopted in faking research for that it is much easier to score than the bias index (Tonkovic, Galic and Jerneic, 2011). Moreover, its validity as faking measures is not reduced by its straightforwardness (Bing et al., 2011). As such, the present study adopts the foil index to measure faking, following Bing et al.'s recommendation. In the following text, the un-

capitalized "foils" refers to the foil items, while the capitalized "OCQ-Foils" refer to the foil index.

Reactions to high-motivated test contexts. The endorsement of foils will be significantly increased under the high-motivated context (Paulhus & Harms, 2004; Bing et al., 2011). In one study, participants' scores inflated substantially under the "fake good" condition (Paulhus et al., 2003). Other researchers found similar results of the bias index when public versus private self-representation demand are salient (Roeder & Paulhus, 2009; Tracy et al., 2009).

According to Paulhus (2011), the results indicate that the bias index is capable of measuring both trait and state self-enhancement. Under low-motivated context, the bias index reflects trait self-enhancement and can be interpreted as "self-deceptive enhancement". Moreover, under high-motivated context, the bias index reflects state self-enhancement and can be interpreted as "impression management". Research on the foil index also found similar results. In the study of Bing et al. (2011), the foil index correlates with SDE but not IM under low motivated context. However, it correlates with both IM and SDE under high-motivated context, and also to a greater extent. In conclusion, the OCQ measure both forms of social desirability, given the appropriate test context.

Correlation with the faking criterion. There is a lack of studies reporting the correlations between OCQ indices and the faking criterion. Also, few studies have looked at the performance of OCQ indices using within-subjects designs under high-low motivated context. More studies are needed in this direction.

Suppression effects. A recent finding about foils showed that the OCQ-Foils serve as a suppressor in the relationship between personality predictor (i.e., achievement striving) and the performance (i.e., GPA) (Bing et al., 2011). The study was a between subjects design including

a motivated condition and a control condition. The suppression effect of OCQ-Foils was only observed in the motivated condition, but not in the control condition. In the study, the suppression effect was tested on several faking measures including IM, SDE, and OCQ-Foils. Among these measures, only OCQ-Foils showed a significant suppression effect under faking condition. Neither IM nor SDE served a suppressor under either honest or faking conditions. As mentioned before, most studies failed to find a suppression effect using SDR measures of faking (Ellingson, Sackett, & Hough, 1999; Hough & Oswald, 2008). The results indicate that OCQ-Foils seem to be a promising measure for faking. To my knowledge, the Bing et al.'s article is the only study that found a suppression effect of OCQ-Foils in the field of IO psychology. More work is clearly needed in this direction.

Moderation effect. There hasn't been any study reporting a moderation effect of OCQ indices upon the criterion-related validity of personalities. However, given the similarity between the bias index and SDE, it is possible that the OCQ-Foils would moderate the criterion-related validity of some personality measures. More study will be needed on this topic.

Issues and comments--similarity and differences between OCQ and BS inventories. According to Paulhus (2011), the OCQ and the BS inventory both can be traced back the work of Raubenheimer (1925). In Raubenheimer's study, participants were provided with a list composed of the names of books. Among the 25 book names, ten of them were foils. For example, the name "Robinson Crusoe" was a true item, which comes from a genuine book. "The Prize-Fighters Story" was not a name of book and thus was a foil item.

In Raubenheimer's study, participants were asked to check which books they had read from the list. This requirement obtained information from two aspects (i.e., past experience, accuracy). The past experience refers to the fact that participants only report the books they have

read. The accuracy refers to whether participants were able to distinguish the real names of the books from the false ones. Based on Raubenheimer's work, the BS inventories and OCT focus differently on these two aspects. The BS items are explicit statements of past experience. They target participants' recall of past behaviors. The OCQ items are terms related to a specific subject, which targets on individual's ability to distinguish information accurately.

The differences and similarities between OCQ and BS inventories have also been indicated by empirical studies. Bing et al. administrated a reduced OCQ, a BS inventory, and self-and-peer reported Big-Five personalities to a sample of job incumbents under low motivated context. The BS and OCQ-Foils showed differential correlations with Big-Five personality traits. The OCQ-Foils showed positive correlations self-rated extraversion and openness as well as negative correlation with peer-rated agreeableness. The BS scores showed positive correlation with self-rated extraversion, and was negatively correlated with self-rated agreeableness and peer-rated conscientiousness. But how OCQ-Foils would act differently from the BS under motivated contexts? More studies are needed to further explain this topic. In the present study, I will make comparison between BS inventory and OCQ in two aspects. To begin with, I will explore whether the reliabilities of OCQ foils would change in a way that is similar to the BS inventory. Specifically, under low-motivated condition, the OCQ foils will show very low reliability; while under high-motivated condition, the OCQ foils will show high reliability. I will also explore whether the BS inventory and the OCQ-Foils will show similar correlational patterns with the faking criterion, and whether they will show similar suppression or moderation effects upon the criterion-related validity of personalities.

Overview of Current Study

The purpose of the present study is to simultaneously examine and compare validity of four faking measures (i.e., IM, SDE, BS inventory and OCQ-Foils). Those measures are administrated under both honest condition and motivated condition. The study examines whether and the extent to which these four faking scores (a) react to the faking context; (b) correlate with residual personality scores in a within-subjects design, (c) suppress personality validities, and (d) moderate personality validities.

The present study is aimed to fill a gap in the faking literature. It does so in two important ways. Firstly, the present study represents the most comprehensive investigation of faking measures thus far, in terms of number of faking measures and number of validity issues simultaneously examined. Secondly, the present study developed a BS inventory that contains more bogus items (i.e., 8 bogus items) than previous BS inventories (see Levashina et al., 2009; Dwight & Donovan, 2003). In the present study, I developed a new BS inventory aimed at college graduates applying for entry-level job positions. This BS inventory contains 25 items including 8 bogus items. The validation of this BS inventory is also represents a contribution to the faking literature.

The study is designed in two stages (i.e., an online baseline assessment, and a lab experiment). In stage one, potential participants complete an online survey. The survey contains half of the personality items (Form A) and serves as the baseline assessment. In stage two, participants coming to the lab are randomly assigned into either a motivated condition or a control condition. Participants from both conditions complete a same set of surveys in the lab environment. These surveys contain the other half of personality items (Form B) as well as the four faking measures (i.e., IM, SDE, BS and OCQ-Foils). Participants in the motivated condition are led to believe that they can get a potential job offer by completing the surveys.

Participants in the control condition are asked to complete the surveys for research purpose only. Participants' cumulative GPA is collected as the performance criterion.

The design of the study serves several purposes. In general, the design enables me to observe the occurrence of faking in both between-subjects and within-subjects manner. To begin with, the experiment creates a high-motivated test situation using the deception technique, which is more close to real-life application situation than using faking instructions. Moreover, incorporating a control condition at Time 2 serves the purpose to verify the success of such an experimental manipulation. Moreover, the control condition is necessary in order to compare the faking measures' responses to the high-motivated test context against a low motivated test context. Finally, the two-stage design makes it possible to observe the faking criterion for the motivated group through the within subjects score change across Time 1 and Time 2.

Hypotheses

Manipulation check. Assume that the experimental manipulation is successful. At least some participants in the motivated condition are motivated to get into the paid summer internship program. As a result, participants of the motivated condition should increase their personality scores substantially from Time 1 to Time 2.

The test context for the control condition also will change from Time 1 to Time 2. To begin with, for the participants of the control condition, the test setting change from a take-home, online test (i.e., Time 1) to a test in the laboratory setting, in which they will take the test with 2-4 people (i.e., Time 2). The social-facilitation effect brought by the laboratory setting in Time 2 can increase the test scores of individuals at the Time 2 control condition. Moreover, participants of both conditions will learn from the informed consent that they will be entered to a drawing for ten $50 cash reward. The mention of monetary incentives will also increase the test

motivation for participants in the control condition. As a result, participants of the control condition would also score higher on personalities at Time 2, comparing with their scores at Time 1.

Despite of the change in test settings, the score change from Time 1 to Time 2 in the control condition should be much smaller than that of the experimental condition. Therefore, I expect to observe a Time-by-Condition interaction effect upon the test scores, such that the score increase from Time 1 to Time 2 in the motivated condition will be significantly higher than that of the control condition. Moreover, there should be significant between-group differences. At Time 2, participants in the motivated condition should score much higher on personality tests, comparing with the control condition. Such a moderation effect has been previously observed by Bing, LeBreton, Davison, Migetz, and James (2007).

Examining the validities of faking measures. Hypotheses 1-6 are aimed to compare and examination the validities of faking measures (i.e., IM, SDE, BS, and OCQ-Foils). The validities of faking measures are examined through their reactions to faking contexts (i.e., hypothesis 1-3), correlations with the faking criterion (i.e., hypothesis 4) and their impactions on the criterion-related validity of personalities upon students' GPA (i.e., hypotheses 5 and 6). Hypotheses 4-6 are based on the assumption that faking only occurred in the motivated condition. As a result, hypotheses 4-6 will only be tested in the motivated condition.

Responses to high-motivated test contexts. Hypothesis 1 is aimed to test the mean score change of all faking measures under high-motivated context. I expect that all the faking measures will show a higher mean under Time 2 motivated condition, compared with the control condition.

Hypothesis 2 is aimed to explore the change of correlational patterns among the faking measures under high-motivated condition, versus the control condition. Specifically, I expect that the faking measures (i.e., IM, SDE, BS and OCQ-Foils) will generate higher correlation with each other at Time 2 motivated condition (i.e., hypothesis 2-a), compared with the control condition.

Hypothesis 3 is aimed to validate the self-developed BS inventory through the change of reliabilities. The success of the BS inventory not only lies in the change of mean scores and correlational patterns (i.e., as has been addressed in hypothesis1). Furthermore, the validity of BS inventory is indicated by change in reliabilities. Specifically, in Time 2 motivated condition, the BS inventory should show high internal consistency. In contrast, the bogus items of BS inventory should demonstrate extremely low reliabilities in Time 2 control condition. The change in reliabilities leads to the hypothesis 2. The hypotheses 1-3 are stated as below.

Hypothesis 1. There should be significant between-group differences on the mean score of all faking measures at Time 2, such that the means of IM, SDE, BS inventory and OCQ-Foils of the motivated condition will be significantly higher than that of the control condition.

Hypothesis 2. a) The IM and SDE will generate higher correlation with each other at Time 2 motivated condition, compared with the control condition. b) The correlation between BS-inventory and IM/SDE will be higher under Time 2 experimental condition, compared with the control condition. c) The correlation between OCQ-Foils and IM/SDE will be higher under Time 2 experimental condition, compared with the control condition.

Hypothesis 3. The bogus items of the BS inventory will exhibit very low internal consistency in the control condition; but will exhibit high internal consistency in the motivated condition.

Aside from the above hypotheses, I also have a few research questions aimed to address the issues mentioned in the issues & comments section of each faking measures. In order to clarify the substance vs. style issue of IM/SDE scales, I will explore to what extent IM/SDE are correlated with personality traits, under both high-and-low-motivated test conditions (see p.18, the IM/SDE section). Moreover, I will explore the correlation between the BS inventory and participants' cognitive abilities (i.e., ACT scores) and academia performance (i.e., cumulative GPA) under Time 2 motivated condition (see p.23, the BS inventory section). In addition, in order to compare the similarities between BS inventory and OCQ, I will also explore the change of internal consistencies of the foil items of OCQ (see the OCQ section, p. 29; this issue is also addressed with the exploratory question 4, p. 35). I suspect the internal consistencies of foils will be higher at Time 2 motivated condition, comparing with the control condition. The exploratory questions 1-3 are stated below.

Exploratory question 1: To what extent will IM/SDE correlate with personality traits, under both high-and-low-motivated test contexts?

Exploratory question 2: Will BS inventory correlate with cognitive abilities and academia performance (i.e., indicated by participants ACT, GPA scores) under Time 2 motivated condition?

Exploratory question 3: Will the reliabilities of OCQ-foils in the Time 2 motivated condition be higher than in the control condition?

Correlation with the faking criterion. Hypothesis 4 is aimed to the correlation between faking measures and the faking criterion. To assume that all faking measures are valid, I propose that all faking measures should achieve significant correlations with the faking criterion. Failure

to demonstrate significant correlations with the faking criterion implies lack of validities or do not work under the lab manipulation.

Hypothesis 4. Under the motivated condition, bogus statement items, over-claiming foils, impression management items, and self-deceptive enhancement items will be positively correlated with the within-subjects score changes of personality traits.

Suppression effects. Hypothesis 5 is aimed to the suppression effects of faking measures upon the criterion-related validity of personalities. The hypothesis assumes that all faking measures will demonstrate such a suppression effect. The emergence of correlations with the faking criterion is more important than the suppression effects. A faking measure can still be valid even if it shows no suppression effects, under the condition that it retains a significant correlation with the faking criterion (i.e., as hypothesis 4 indicates).

Hypothesis 5. Under high motivated condition, IM, SDE, BS and OCQ-Foils will show suppression effects on criterion-related validity of personality scores (i.e., Big-Five and school-specific conscientiousness), where the criteria are GPA.

Moderation effects. Hypothesis 6 is aimed to the moderation effects of faking measures upon the criterion-related validity of personalities. The hypothesis assumes that all faking measures will demonstrate a significant moderation effect. The emergence of correlations with the faking criterion is more important than the moderation effects. A faking measure can still be valid even if it shows no moderation effects, under the condition that it retains a significant correlation with the faking criterion (i.e., as hypothesis 4 indicates).

Hypothesis 6. Under high motivated condition, there should be interactive effects between IM, SDE, BS, OCQ-Foils scores and personality scores in predicting various criteria

such that for participants with a high IM, SDE, BS, OCQ-Foils score, the personality-criteria correlations should be lower than participants with a low IM, SDE, BS, OCQ-Foils.

Comparing the validities of faking measures. On the basis of hypotheses 4-6 the validities of the faking measures will be compared. I will compare the size of correlation coefficients of IM, SDE, BS, and OCQ-Foils towards the residual scores of personalities (i.e., IPIP-100, Goldberg et al., 2006; school-specific conscientiousness, Schmit, Ryan, Stierwalt, & Powell, 1995). I will also compare the amount of suppression/moderation effects that the four faking measures show upon the criterion-related validity of personalities. The higher correlation that a faking measure demonstrate with the faking criterion, the more valid it is; and the more a faking measure showed suppression or moderation effects upon the criterion-related validity of personalities, the more valid it is.

Exploratory Question 4: Will IM, SDE, BS, and OCQ-Foils show similar correlation coefficients with the residual scores of personalities? Will IM, SDE, BS, and OCQ-Foils show similar suppression/moderation effects upon the criterion-related validity of personalities?

Exploring the validities of faking measures under the control condition. Faking research conducted under a low-motivated context typically found a non-significant correlation between faking measures and the faking criterion (e.g., Griffith et al., 2006). These studies also tend to find negative results towards the moderation/suppression effects of faking measures (e.g., Ganster at al., 1983; Borkenau & Ostendorf, 1992). In the present study, I will explore whether the four faking measures would correlate with the faking criterion under the control condition, and whether these faking measures would demonstrate suppression and/or moderation effects under the control condition. My suspect is, in the control condition, the faking measures will no longer correlate with the faking criterion, and neither would they show suppression or

moderation effects upon the criterion-related validity of personalities. The explanatory questions for the control condition are stated as below.

Exploratory Question 5: Will IM, SDE, BS, and OCQ-Foils show correlation with the faking criterion at Time 2 control condition?

Exploratory Question 6: Will IM, SDE, BS, and OCQ-Foils show the suppression/moderation effects upon the criterion-related validity of personalities at Time 2 control condition?

Method

 Sample and Procedure

Participants are undergraduate students recruited through the SONA system operated by the Psychology Department of Auburn University. Participants are at least 19 years old.

The study is a two-stage study. In stage one, students complete an online Mass Screening questionnaire through the SONA system. The survey contains half of IPIP-100 personality items (Form A, 50 items; Goldberg et al., 2006) and half of school-specific conscientiousness items (Form A, 12 items; Schmit, Ryan, Stierwalt, & Powell, 1995). Only those who have completed the mass-screening questionnaire are eligible for signing up for the lab study.

In stage two, participants who sign up for the lab study are randomly assigned into either a motivated condition or a control condition. Participants in both conditions are asked to complete a diversion task framed as the resume content analysis task. After that, participants in the experimental condition are deceived. They are told the cover story that researchers are working with a company to develop a test for recruiting students into a paid summer-internship program. The Company needs a group of college students to tryout the test. In order to thank students for their participation in the study, the Company provides two additional benefits. Firstly, the Company has agreed to sponsor a drawing for ten $50 cash rewards among all participants, which will occur by the end of data collection. Secondly, the Company also provides a chance for students to apply for the paid internship program by participating in the research. If students indicate their interests in the internship program by selecting "yes" at the end of the assessment, the Company will view their responses more carefully and give them early considerations. Students are also suggested to pay a close attention to an introduction at the first page of the survey, which contains more detailed information of the Company and the

program. At the same time, participants in the control condition are told to complete the same set of survey with the same format for research purposes only. On the first page of survey, participants of the control condition will also read a survey introduction, which reassures that the information will be kept confidential and instructs participants to respond as honestly and possible.

After the experimental manipulation, participants in both conditions will complete a set of surveys. The surveys include (a) another half of IPIP-100 (Form B) and school-specific conscientiousness items (Form B), (b) a 10-item impression management inventory, (c) a 10-item self-deceptive enhancement inventory, (d) a 25-item bogus statement inventory, and (e) a 25-item short-form of over-claiming questionnaire. The items of a), b), and c) are mixed together and are constructed into a "personality assessment", followed by the items of (d) as a "task experience assessment", and in-turn followed by the items of (e) as a "general knowledge assessment".

All participants are debriefed after they complete the surveys, in which the deception is explained and students are given a chance to withdraw without losing any benefits (i.e., SONA credits, and the drawing). In order to collect a performance criterion, participants' cumulative GPA and ACT/or SAT scores are obtained, with their permission, through the University Registrar. The cumulative GPA is used as the performance criterion, whereas ACT or SAT scores are controlled in subsequent analyses. All the items are available at Appendix A. A detailed deception script for the present study is attached in appendix B, whereas the survey instructions for both the motivated condition and the control condition are included in Appendix C. The debriefing script for both conditions can be found in Appendix D.

Both the lab procedure and the deception script used in the present study are based on Ellingson, Heggestad, and Makarius' (2012)'s work. The present study shares similarities with Elligson et al. study (i.e., deception technique, lab procedures). However, in order to fit our participants, I make a few changes in the deception script. Ellingson et al. used a faked leadership development program to motivate their participants from the college of business. In the present study, I replaced the "leadership development program" with a "paid summer-internship program" to attract our participants from the college of liberal arts.

The present study is aimed to investigate hypotheses and research questions that are inherently different from Ellingson et al.'s study. The present study focus on the validation and comparison of four faking measures under high motivated context, whereas the study of Ellingson et al.'s was aimed to test warning's impact upon applicants faking and the underline mechanism. Therefore, the adaptation of deception technique from a published article does not reduce the innovation and purposes of the present study. My student plan and procedure was approved by the Institution Review Board of Auburn University.

Measures

Personality measures.

Big-Five personality. The Big-Five personality traits are measured by Goldenberg IPIP-100 items (International Personality Item Pool, Goldberg et al., 2006), which is a popular measure for the Big-Five personalities. The IPIP-100 has been split into two 50-item forms (i.e., Form A, Form B). Half of the items (i.e., the Form A) are used in the baseline assessment, while another half items (i.e., Form B) are used for the lab experiment. Each form taps in all the five factors and measure each factor with 10 items. For each question, participants rate their

agreement with the statements on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). The way items were split was based on Ellingson et al. (2012).

School specific conscientiousness. The school specific conscientiousness items come from the work of Schmit et al. (1995), in which the authors re-wrote the conscientiousness dimension of NEO-PI-R (Costa & McCrae, 1992) to be suitable for the school context. The school-specific conscientiousness items have been validated (Schmit et al., 1995; Bing, Whanger, Davision & Vanhook, 2004). The latter study shows that the school-specific items have incremental validity above and beyond general personality measures for academic performance.

The school specific contentiousness includes a total of 24 items. It has three dimensions (i.e., achievement striving, self-disciplinary and competitiveness); each dimension is measured by 8 items. The school specific contentiousness items are randomly split into half and reframed into form A and form B (Bing et al., 2004). Each form contains 12 items which measuring each sub-dimension with 4 items. For each question, participants rate their agreement with the statements on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Half of the school-specific conscientiousness, the form A, is used for the baseline assessment. Another half of the school-specific conscientiousness, the form B, is used for lab assessment.

All personality surveys are self-report. The residual scores of personalities are measured in the lab experiment over the baseline assessment in the experimental condition, which represents an objective indicator of faking (i.e., the faking criterion).

Faking measures.

IM and SDE. A measure of IM and SDE are used based on the two-factor model of social desirability (Paulhus, 1984). The measure contains 10 items for IM and another 10 items

for SDE. A sample item for IM is ''I always apologize to others for my mistakes.'', while a sample item for SDE is ''Life is a strain for me most of the time'' (reverse-coded). Responses to each item are coded on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

In this measure, 7 items of IM and 5 items of SDE are based on a same measure from Bing et al. (2011). Bing chose these items from Paulhus (1984) and Crown and Marlowe (1960). The rest of the IM/SDE items are selected from the BIDR-6 (Paulhus, 1988). I decide to use a reduced version of BIDR in the present study because the full version (i.e., 40-item) is too lengthy and time consuming to administrate under the lab condition.

Bogus statement inventory. The bogus statement is assessed with a self-developed BS inventory. The BS inventory is composed of 25 items (i.e., 8 bogus items and 17 genuine items). I developed the BS inventory for two main reasons. Firstly, there has been missing a BS inventory for college graduates applying for general entry level jobs. Secondly, I want to increase the amount and proportion of bogus items in a BS inventory. I and another graduate student majoring in I/O psychology served as subject matter experts (SMEs). We came up with an item pool of BS inventory for both genuine items and bogus items. The items used in the BS inventory were finalized through the group discussion among the two SMEs, the advisor of the present author and two undergraduate research assistants majoring in psychology. The items are tasks tapping in multiple areas including research skills, computer skills, basic clerical skills and interpersonal skills. Among those, three out of six basic research skills items are bogus items, three out of six computer skills are bogus items, zero out of the eight basic clerical skills is bogus items and two out of three interpersonal skills are bogus items. One item from Dwight and Donovan's (2003) was included in the present BS inventory, which is "Utilize Murray-Web

system to locate unpublished research articles" (a research skill BS item). For each task, respondent indicate their frequency of engaging in the task in the past 12 month on a four-point Likert-scale ranging from 1 (Never) to 4 (Often).

Over-claiming questionnaire. Over-claiming is measured with a reduced 25-item short form (OCSF25) which is developed and validated by Bing et al. (2011). The OCSF-25 scale is adapted from the OCQ-150 (Paulhus et al., 2001). The OCSF 25 includes 8 foils and 17 genuine items. All items are selected from the OCQ-150. For each question, participants rate their familiarity with the terms from a scale ranging from 1 (Never heard of it) to 5 (Very familiar).

Dependent variables. Students' GPA, ACT and /or SAT scores are obtained from the University Register by the end of Semester of fall 2013. The GPA is used as the academic performance criterion. The ACT/SAT scores are controlled in subsequent analysis. The study chose the criterion based on Bing et al. (2011), who also used GPA as the performance criterion.

Analytical Strategies

All manipulation check questions regarding the mean score change of personality tests will be tested with one-way MANOVA. The Time-by-Condition moderation effect will be tested with mixed MANOVA.

The mean score changes of faking measures between Time 2 motivated condition and control condition will be tested with one-way ANOVA (i.e., hypothesis 1). The comparisons of correlation coefficients (i.e., hypothesis 2-a, 2-b, and 2-c) will be tested with the Fisher z-transformation for the Pearson correlation test (Fisher, 1915). The comparison of reliabilities (i.e., hypothesis 3) will be tested in the "cocorn" package in R (Diedenhofe, 2013), which was developed based on the theoretical work of Charter and Feldt (1996) and Feldt (1969, 1980).

The significance of correlation between faking measures and the faking criterion (i.e., hypothesis 4) will be tested with the Pearson bivariate correlational test.

The suppression effects (i.e., hypothesis 5) of faking measures will be tested with 3-steps linear regression method. In step one, the dependent variable (i.e., GPA scores) will be regressed on the control variable (i.e., ACT scores). In step two, the personality predictor will be entered into the regression equation. In step three, the faking measure, which serves as a suppressor, will be entered into the regression equation. A significant suppression effect will be established if a) the R^2 change from step-two to step-three is significant; b) the regression coefficient for the personality predictor increase substantially from step two to step three; and c) the suppressor variable shows a significant negative correlation with the dependent variable at step three. See also Bing et al. (2011) for the above procedure.

The moderation effects (i.e., hypothesis 6) of faking measures will be tested with linear regression method. In step one, the dependent variable (i.e., GPA) will be regressed onto the personality predictor and the moderator, with no interaction term. In step two, the interaction term (i.e., the product term of the personality predictor and the faking moderator) will enter the regression equation, and significance tests will be examined. If the interaction term is significant, a simple slope test will be conducted, following the methods and steps introduced by Preacher, Curran, and Bauer (2006).

The tests for the explanatory questions are similar to the hypotheses. The explanatory question 1, 2 and 4 will be tested with Pearson correlation tests. The explanatory question 3 will be tested in R with the "cocorn" package (Diedenhofe, 2013). The explanatory question 5 and 6 will follow the procedures of testing suppression and moderation effects, as described above.

Results

Participants were undergraduate students recruited through the SONA system operated by the Psychology Department of Auburn University. There was a mass-screening study posted in the SONA system, which included our baseline measures (IPIP-100, Form A and Bing's school-specific Conscientiousness items, Form A) and other researchers' measures (i.e., Time1). The mass-screening study was offered throughout the fall 2013 semester, and only those who had completed the mass screenings study were eligible for the lab study (i.e., Time 2).

Descriptive Statistics

Sample size

A total of 161 students participated in the lab experiment of the study (i.e., Time 2) at Auburn University during fall 2013. A total of 86 participants were tested in the motivated condition, and 75 in the control condition. Among all participants, 119 of them also participated in the baseline assessment through the mass-screening survey (i.e., Time 1). A substantial percentage of subjects (i.e., 26%) did not complete the baseline assessment, but somehow got into the Time 2 of the study. For participants who completed both Time 1 and Time 2, 61 were in the motivated condition, and 58 were in the control condition (see Table 1 for the means, standard deviations, and reliabilities of personality scores).

Equating

In the present study, participants' personalities were measured twice using the parallel forms (i.e., form A and form B; see Ellingson, 2012). To control for the potential differences between form the two parallel forms, an equating process was conducted before data analyzing. In the present study, participants of the motivated condition completed the personality tests under

different test condition (i.e., honest vs. motivated).   Participants of the control condition

completed both tests under the same test condition—the honest condition.   Participants of the

Table 1

Descriptive Statistics for Personality Scores (Before Equating)

Scale	Motivated [a]						Control [b]					
	Time 1			Time 2			Time 1			Time 2		
	M	*SD*	*α*	*M*	*SD*	*α*	*M*	*SD*	*α*	*M*	*SD*	*α*
Agreeableness	5.30	.82	.78	5.93	.67	.83	5.17	.86	.79	5.76	.76	.85
Conscientiousness	4.49	1.05	.85	5.37	.83	.83	4.36	1.09	.88	4.71	.86	.81
Emotional Stability	4.53	.99	.81	4.66	.98	.87	4.68	.91	.79	4.49	1.09	.89
Extroversion	4.44	.94	.81	5.29	.88	.85	4.54	.84	.79	5.28	.97	.88
Openness	4.72	.77	.76	5.28	.76	.79	4.66	.75	.75	4.96	.83	.77
Achievement Striving	4.29	1.07	.61	5.34	1.14	.76	3.97	1.13	.68	4.86	1.25	.65
Competitiveness	4.63	1.16	.70	5.41	.96	.67	4.62	1.02	.56	5.11	1.00	.45
Self-disciplinary	4.44	1.16	.70	5.33	1.20	.84	4.56	.99	.53	4.84	.53	.85
Total score of school conscientiousness	4.45	.95	.83	5.36	.99	.86	4.38	.91	.89	4.92	1.09	.88

Note. Participants completed both Time1 and Time 2 were included, [a]*n* =61.[b]*n*=58.

47

control condition did not change the test motivation at Time 2, it is assumed that any differences between Time 1 and Time 2 on the personality scores of the control condition were due to random errors caused by item difference only. To ensure that there is no differences between the two parallel forms, all participants' personality scores at Time 2 were adjusted by subtracting the differences between the mean scores of personalities between Time 2 and Time 1 of the control condition (i.e., Equated score=Raw score- (Mean $_{\text{Time 2 Control}}$- Mean $_{\text{Time 1 Control}}$)). The adjusting process was referred to as equating.

The equating not only eliminates the item-inequity of parallel forms of personality tests, but also eliminates other differences across Time 1 and Time 2 as well. At Time 2, all participants completed the questionnaires under the lab environment in groups of two to six people under the presence of an experimenter. The presence of peers and experimenter can trigger the social facilitation effect, which would increase participants' testing motivation and make them perform better in the test. Moreover, participants of both conditions learned about the potential monetary incentive brought by participating in the study (i.e., $50 cash reward drawing). Knowledge of the monetary incentives can also increase the test motivation of participants at Time 2. By subtracting the means differences from Time 2 to Time 1 in the control condition, the differences brought by these two factors would also be eliminated. Any difference between Time 1 and Time 2 would be merely due to the experimental manipulation itself. All personality scores used in subsequent analyses were equated. The means, standard deviations, and reliabilities of personalities after equating were available in Table 2.

Table 2

Descriptive Statistics for Equated Personality Scores at Time 2

Scale	Motivated [a]			Control [b]		
	M	*SD*	*α*	*M*	*SD*	*α*
Agreeableness	5.44	.66	.83	5.19	.72	.84
Conscientiousness	5.08	.83	.82	4.43	.94	.84
Emotional Stability	4.61	1.00	.87	4.53	1.13	.90
Extraversion	4.96	.88	.85	4.62	.96	.88
Openness	4.97	.83	.81	4.67	.80	.76
Achievement Striving	4.58	1.12	.71	4.11	1.26	.70
Competitiveness	5.02	.98	.68	4.72	1.02	.50
Self-discipline	5.21	1.18	.83	4.66	1.32	.85
All-school specific Conscientiousness	4.95	1.01	.89	4.50	1.10	.89

Note. All participants were included, [a]*n* =86.[b]*n*=75.

Demographics

The demographic information (i.e., age, gender, race and ethnicity) was collected with the mass-screening survey. For the students that completed the mass-screening survey and participated in the lab experiment, 70.6% were female, and the mean age was 19.76, ranging from 19 to 25. 86.6% of these participants were Caucasians, followed by 6.7% Africa Americans, 4.2% Asian and Pacific islanders, and 1.7% Hispanic. There was no significant difference in age, gender, race and ethnicity between the motivated condition and the control condition.

GPA

Participants' cumulative GPA as well as ACT or SAT scores were obtained from the University Registrar right after the end of fall 2013 semester. The GPA was obtained for 139 participants, and the ACT or SAT scores were obtained for 136 participants. The full academic information (i.e., both a GPA and a standard test score, either ACT or SAT) was obtained for 135 participants, with 75 in the motivated condition and 60 in the control condition.

Since majority students had ACT rather than SAT scores, I translated the SAT scores into ACT scores according based on the methods described on the ACT official website ("Compwere ACT & SAT Scores", 2008). There was no significant difference in GPA ($t = -.66$, $p > .10$) and ACT ($t = -1.58$, $p > .10$) between the motivated condition and the control condition. The mean of GPA in the current sample is 3.01, and the mean for ACT is 25.38.

Interest in the paid summer internship program

In the motivated condition, there was a question at the very end of the questionnaire asking whether participants would like to be contacted/receive further information about the paid summer internship program. If participants wished to be further contacted for joining the

program, they could select "Yes"; otherwise they would select "No". Among the 86 students who participated in Time 2 motivated condition, 25 selected "Yes", 60 selected "No", and one student left the question blank (i.e., it is a missing value). Given the fact that less than 30% students indicated their interest in the paid summer internship program, the level of test motivation in the present study was moderate.

Scoring method of the BS inventory and OCQ-Foils

There are two ways to score the BS inventory: a dichotomous method, and a continuous method. Fan et al. (2012), Levashina et al. (2009) and Pannone (1984) used a dichotomous approach. In this method, a bogus item will simply be scored as endorsed (i.e., counted as 1) and not endorsed (i.e., counted as 0). Any indication of familiarity with a bogus item will be counted as endorsing, regardless of the extent of familiarity, and participants could only avoid endorsing a bogus item by claim a completely unfamiliarity with the item (i.e., by choosing "Never"). In contrast, Dwight and Donavan (2003) used a continuous approach. In the continuous scoring method, the bogus items were viewed as a regular personality items, and the score on bogus item is indicated by the value of liker-scale associating with the option chosen (i.e., a scale of frequency range from 1 to 4, 1 for "Never", and 4 for "Often").

In the present study, the BS inventory was scored using both methods. The dichotomous BS scores were calculated by the sum of eight dichotomously scored bogus items (i.e., each item was scored at 0 or 1; total score ranged from 0 to 8). The dichotomous BS score indicated the number of endorsed bogus items. The continuous BS scores were calculated as the average scores across the eight continuous bogus items (i.e., range from 1 to 4, 1 for "Never", and 4 for "Often"). The continuous BS score indicated the average frequency of engaging in the bogus tasks. I also scored OCQ-Foils both dichotomously and continuously, similar to the method of

the BS inventory, for comparison purposes.   Scoring BS and OCQ-Foils into both dichotomous

and continuous scores generated a total of six faking variables (i.e., IM, SDE, BS scored

dichotomously, BS scored continuously, OCQ-Foils scored dichotomously, and OCQ-Foils

scored continuously).

The self-developed BS inventory

The present study incorporates a self-developed BS inventory. It is necessary to validate

the BS inventory before any further analysis. The BS inventory consisted of eight bogus items.

The means and SD for the items are available in Table 3.   The descriptive statistics indicated

that, in general, the mean scores on the bogus items were low. The scores on bogus items were

low for both motivated condition and control condition, and for both scoring methods.   Most

items lingered around the bottom line of the rating scales (i.e., for dichotomized scoring, around

0; for continuous scoring, around 1).

Since the average scores on the BS items were low, it was possible that the bogus items

would not be able to differentiate the two conditions. In order to further examine this possibility,

I conducted two one-way MANOVA tests to compare the means of bogus items across the lab

conditions (i.e., motivated, and control). Results indicated that the MANOVA tests were not

significant, $F (1, 159) =.83$, $p=.57$ for dichotomously scored bogus items, and $F (1, 159) =.99$,

$p=.40$, for continuously scored bogus items. The univariate results are available in Table 4.

Results indicated that among the eight bogus items, only one (i.e., the BS 5, for dichotomous

scores) or two (i.e., the BS 5 and BS 17, for continuous scores) items showed significant between

group effects. In general, the bogus items failed to show enough differentiation between the

motivated condition and the control condition.

52

Table 3

Descriptive statistics of the self-developed BS inventory

		Dichotomously scored				Continuously scored			
		Motivated [a]		Control [b]		Motivated [a]		Control [b]	
		M	SD	M	SD	M	SD	M	SD
BS 1	Utilize the Murray-Web system to locate unpublished research articles	.13	.34	.12	.33	1.13	.37	1.16	.55
BS 5	Organize files using the FolderPro Filing Scheme	.15	.36	.03	.18	1.24	.65	1.09	.44
BS 8	Use Johnson's Dyadic Approach of avoiding conflict in work teams.	.18	.39	.10	.31	1.17	.41	1.13	.38
BS 10	Assess the feasibility of a business project using the Expert Assessment	.03	.18	.02	.13	1.03	.24	1.01	.12
BS 15	Use Rastmen's Power theory to motivate others in your team.	.10	.30	.03	.18	1.13	.48	1.07	.30
BS 17	Create slideshow using Picslide program	.19	.40	.05	.22	1.22	.58	1.09	.34

BS 21	Use the silvernman leadership skill to distribute teamwork appropriately	.15	.36	.16	.37	1.20	.48	1.21	.55
BS 24	Operate a Nephogram projector	.05	.22	.02	.13	1.06	.24	1.03	.16

Note. Participants completed both Time1 and Time 2 were included, [a]n =61.[b]n=58.

Table 4

Item-level analysis of the self-developed BS inventory

		Dichotomously scored				Continuously scored			
		Condition [a]		Motivation [b]		Condition [a]		Motivation [b]	
		F	η^2	F	η^2	F	η^2	F	η^2
BS 1	Utilize the Murray-Web system to locate unpublished research articles	.04	.00	1.22	.02	.19	.00	.50	.01
BS 5	Organize files using the FolderPro Filing Scheme	4.11	.03*	2.53	.03†	2.89	.02†	1.53	.02
BS 8	Use Johnson's Dyadic Approach of avoiding conflict in work teams.	.59	.00	.60	.01	.43	.00	.26	.00
BS 10	Assess the feasibility of a business project using the Expert Assessment	.21	.00	3.25	.04*	.50	.00	3.81	.05*
BS 15	Use Rastmen's Power theory to motivate others in your team.	.49	.00	1.94	.02	.91	.01	2.96	.04†
BS 17	Create slideshow using Picslide program	2.53	.02	5.47	.07*	2.79	.02†	8.75	.10**
BS 21	Use the Silvernman leadership skill to distribute teamwork appropriately	.00	.00	.73	.01	.04	.00	1.74	.02

55

BS 24	Operate a Nephogram projector	.95	.01	5.04	.06[*]	.95	.01	5.04	.06[*]

Note. df [a] = (1, 159), df [b] = (2, 157). †$p < .10$ *$p<0.05$ ** $p<0.01$.

The results that only one or two bogus items showed significant between-group effect at Time 2 were not as perfect. However, this result can be explained by another factor. As earlier mentioned, only one third participants of the motivated condition indicated their interest in the paid summer internship program (i.e., 25 out of 86), while majority participants of the motivated condition did not (i.e., 61 out of 86). In fact, the lack of differentiation of bogus items may be a consequence of the fact that majority of the participants in the motivated condition were not motivated enough to inflate their scores on the BS inventory. The bogus items may show mean differences on levels of test motivation, as identified by whether participants showed an interest in the paid internship program.

To verify the speculation, I compared the mean scores of bogus items again on a three-level grouping variable using two additional MANOVA tests. Specifically, participants of the motivated condition who also indicated an interest in the program had the highest test motivation. The rest of participants of the motivated condition, who did not indicate an interest in the internship program, had the second highest test motivation. Participants of the control condition had the lowest test motivation. Results indicated that the two MANOVA tests were significant, $F (2, 157) =2.09$, $p<.01$ for dichotomous scores, and $F (2, 157) =2.73$, $p<.01$, for continuous scores. The results indicated that, consistent with my speculation, the bogus items were able to differentiate different levels of test motivation.

The univariate results are included in Table 4. In a closer examination of the univariate results, I noticed that the eight bogus items showed differential effects upon the levels of test motivation. At one hand, four out of eight bogus items showed significant between-subjects effects upon the levels of test motivation (i.e., BS 5, BS 10, BS 17, and BS 24, for dichotomous scores; BS 10, BS 15, BS 17, and BS 24, for continuous scores). At the other hand, there were

also some bogus items that did not show any difference on either lab conditions or levels of test motivations. As Table 4 indicates, the BS 1, BS 8 and BS 21 did not show any difference on either lab conditions or test motivations. Moreover, although the BS 15 showed significant mean differences upon levels of test motivation, this effect was only observed for continuous scores, not for dichotomous scores. In general, the BS 1, BS 8, BS 15 and BS 21 failed to differentiate participants upon either lab conditions or test motivations. As a result, these items were dropped from the final version of BS. Only the items that showed significant differentiations (i.e., BS 5, BS 10, BS 17, and BS 24) were retained for further analyses. With the finalized four-item BS inventory, the descriptive statistics of all faking measures were available at Table 5.

Manipulation Check

In the present study, a successful lab manipulation (i.e., the deception technique) would lead to inflated test scores at Time 2 motivated condition. For manipulation check purposes, a one-way MANOVA test was conducted to compare the means of equated personality scores at Time 2 between the lab conditions. The MANOVA test was significant, F (8, 152) =3.19, $p<.01$. The univariate tests results are available in Table 6. Significant between-subjects differences were obtained on all personality dimensions aside from extraversion (p=.56). The results indicated that participants of the motivated condition successfully increased their scores on most personality dimensions, compared to the participants of the control condition. The inflated personality scores indicated that in general, the lap manipulation was successful, and the deception technique worked in the present study.

Table 5

Univariate Tests for equated Personality Scores at Time 2

Scale	Condition	
	F	η^2
Agreeableness	5.46	.03[*]
Conscientiousness	22.24	.12[**]
Emotional Stability	4.04	.02[*]
Extroversion	0.34	.00
Openness	5.44	.03[*]
Achievement Striving	6.25	.04[*]
Competitiveness	3.63	.02[†]
Self-disciplinary	7.77	.05[**]
Total score of school conscientiousness	7.34	.04[**]

Note. Degree of freedom: (1, 117).

[†]$p < .10$ [*]$p<0.05$ [**]$p<0.01$.

Table 6

Descriptive Statistics for the faking measures at Time 2

Scale	Motivated [a]			Control [b]		
	M	*SD*	*α*	*M*	*SD*	*α*
IM	5.03	0.82	0.79	4.3	0.86	0.75
SDE	5.16	0.71	0.71	4.84	0.78	0.7
BS dichotomous	0.43	0.74	0.49	0.17	0.41	0.05
BS continuous	1.14	0.29	0.47	1.06	0.15	-0.03
FOILS dichotomous	3.74	1.92	0.66	2.93	1.75	0.60
FOILS continuous	1.82	0.58	0.79	1.62	0.53	0.73

Note. IM=Impression Management, SDE=Self-Deceptive Enhancement. The BS scored dichotomously represents the total score of four dichotomized BS items (i.e., each item scored at 0 or 1; the total score range from 0 to 4), while the BS scored continuously represents the average score of four continuously scored BS items (i.e., range from 1 to 4; 1=never, 2=a couple of times, 3=quite a number of times, 4=often). [a]*n* =61.[b]*n*=58.

Examining the Validities of Faking Measures

Faking Measures' Responses to High-Motivated Test Context

Mean-score change. Hypothesis 1 looked at the between-subjects effect (i.e., the lab condition) upon the mean scores of the faking measures (i.e., IM, SDE, BS inventory and OCQ-Foils) at Time 2. To test the hypothesis, a one-way MANOVA test was conducted, with the faking measures to be the dependent variables, and the condition (i.e., motivated, control) to be the fixed factor. Note that both the BS inventory and the OCQ-Foils were scored in two methods (i.e., dichotomous, and continuous). There were six dependent variables for the one-way MANOVA test (i.e., IM, SDE, BS dichotomous, BS continuous, OCQ-Foils dichotomous, and OCQ-Foils continuous). The results indicated that in general, there was a significant between-subjects difference on the faking measures, F (6, 154) =4.90, $p<.001$.

The univariate results are available in Table 7. Consistent with the hypothesis 1, participants obtained higher mean scores on all four faking measures (i.e., IM, SDE, BS and OCQ) under the motivated condition, compared to the control condition. The IM obtained the largest effect size (η^2=.13, p<.01), followed by SDE (η^2=.05, p<.01) and dichotomous OCQ-Foils (η^2=.05, p<.01), followed by the BS inventory (i.e., both dichotomous and continuous, η^2=.03, p<.05) and continuous OCQ-Foils (η^2=.03, p<.05). The results indicated that consistent with previous findings, IM is most sensitive to test motivation (i.e., Paulhus et al., 2003; Paulhus, 2002). Participants also increased their scores of other faking measures, but only to a less extent. The results indicated that to more or less extent, all four faking measures reflected increase in test motivation via mean score change. Initially, all faking measures worked under the motivated condition.

Correlational change. Hypothesis 2 looked at the correlational change of the faking measures under high-motivated context. Specifically, hypothesis 2 examined whether BS and OCQ-Foils would correlate higher with IM/SDE under motivated condition, compared to the control condition (i.e., see Table 8 for the zero-order correlations of the above faking measures under both lab conditions).   In order to test the hypothesis 2, fisher z-tests for comparison of Pearson correlations were conducted, following the methods described by Preacher (2002). The fisher z-tests results are available in Table 9. The results indicated that only OCQ-Foils and IM correlated higher under motivated condition, compared to control condition. Aside from this exception, none of the pairs were correlated higher at motivated condition, compared to control condition.   Therefore, only the hypothesis 2-c was partially supported; the hypotheses 2-a and 2-b were not supported. The results indicated that, opposite to the hypothesis, the faking measures did not convergent under the motivated condition. The results may due to that different faking measures captured different aspects of applicant faking under the simulated application situation.

Table 7

Univariate Tests for the Faking Measures at Time 2

Scale	Condition (Time 2)	
	F	η^2
IM	23.39[*] [*]	.13
SDE	9.22[**]	.05
BS dichotomous	5.29[*]	.03
BS continuous	4.99[*]	.03
FOILS dichotomous	7.77[**]	.05
FOILS continuous	4.61[*]	.03

Note. IM=Impression Management, SDE=Self-Deceptive Enhancement. Degree of freedom:

(1, 159). n=160; all participants were included.

Table 8.

Correlation Coefficients of Faking Measures

Scale	IM	SDE	BS dichotomous	BS continuous	FOILS dichotomous	FOILS continuous
IM	-	.51**	.00	.00	.23*	.27*
SDE	.65**	-	.02	.03	-.03	.05
BS dichotomous	-.10	-.07	-	.94**	.49**	.53**
BS continuous	-.04	.00	.90**	-	.44**	.46**
FOILS dichotomous	-.04	.02	.31*	.33*	-	.85**
FOILS continuous	-.04	.05	.25**	.27*	.85**	-

Note. IM=Impression Management, SDE=Self-Deceptive Enhancement. The upper diagonal part represents the correlation coefficients under the motivated condition, while the lower-diagonal represents the correlational coefficients under the control condition. All participants were included, n=86 for the motivated condition, n=75 for the control condition. *p<0.05. ** p<0.01.

Table 9.

Results of Fisher's Test of Correlation Coefficients between Faking Measures

Correlation	r motivated	r control	Fisher's z	p-value (one-tailed)	p-value (two-tailed)
IM-SDE	.51**	.65**	1.31	.10†	0.19
IM-BS dichotomous	.00	-.10	.62	.27	.53
IM-BS continuous	.00	-.04	.25	.40	.80
IM-OCQ dichotomous	.23*	-.04	-1.46	.05†	.09†
IM-OCQ continuous	.27*	-.04	-1.97	.02*	.05*
SDE-BS dichotomous	.02	-.07	.56	.29	.58
SDE-BS continuous	.03	.00	.19	.42	.85
SDE-OCQ dichotomous	-.03	.02	.32	.37	.75
SDE-OCQ continuous	.05	.05	.01	.50	1

Note. IM=Impression Management, SDE=Self-Deceptive Enhancement. n=86 for the motivated

condition, n=75 for the control condition. $^†p < .10$ *p<0.05. ** p<0.01.

Alpha coefficients change for BS inventory. Hypothesis 3 was aimed to validate the

self-developed BS inventory through the change of reliabilities between different lab conditions.

Hypothesis 3 expected that the reliability of the BS inventory (i.e., the alpha coefficients of the

four finalized bogus items) would be higher in the motivated condition, compared to the control

condition. In order to test the hypothesis, W-tests (Feldt, 1969) were conducted in R using the

cocron-package (Diedenhofen, 2013). The results for W-tests are available in Table 10. The

reliabilities of the BS inventory of the motivated were significantly higher than that of the

control condition, $F (85, 74) =1.86$, $p< .01$, for dichotomous scores, $F (85, 74)=1.88$, $p< .01$, for

continuous scores. The results indicated that the self-developed BS inventory was successful

from a perspective of reliability change.

   Correlation between IM/SDE and personalities. Explanation question 1 was aimed to

investigate whether IM/SDE scales would correlate with personality traits, under both motivated

condition and the control condition. To examine this question, I conducted bivariate correlation

analyses between IM, SDE and all personality scores (i.e., IPIP Big-Five and school specific

conscientiousness dimensions). The results of the correlation analyses are available in Table 11.

Under motivated condition, IM and SDE showed moderate to high correlations with all

personalities scores, with r ranged from .32 to.76, $p<.01$. Under the control condition, IM and

SDE also showed significant correlations with most personality scores, with three exceptions.

The correlation between IM and extraversion was not significant, $r=.07$, $p>.05$. The correlation

between IM/SDE and openness were also not significant, $r=.01$ and .14, respectively, both $p>.05$.

The results indicated that consistent with previous literature (i.e., Holden & Passey, 2012), IM

and SDE correlated substantially with personality dimensions, under high and low motivated test

Table 10.

Between-Group Alpha Coefficient Analysis for BS Inventory

Scale	Motivated Condition	Control Condition	*W (two-tailed)*
BS dichotomous	0.49	0.05	1.86**
BS continuous	0.47	-0.03	1.88**

Note. W is a statistical test based on the sampling distribution of coefficient alpha which suggests that the ration (W) is distributed as a central F with dfs (N1-1, N2-1). The degrees of freedom of the above tests were (85, 74).

†*p* < .10, *p<0.05, ** *p*<0.01.

Table 11.

Correlation Coefficients Between Impression Management, Self-Deceptive Enhancement, and Personalities.

Scale	Motivated [a]		Control [b]	
	IM	SDE	IM	SDE
Agreeableness	.56[**]	.40[**]	.51[**]	.45[**]
Conscientiousness	.55[**]	.43[**]	.44[**]	.32[**]
Emotional Stability	.52[**]	.76[**]	.57[**]	.71[**]
Extroversion	.34[**]	.39[**]	0.07	.32[**]
Openness	.56[**]	.57[**]	0.01	0.14
Achievement Striving	.36[**]	.32[**]	.48[**]	.49[**]
Competitiveness	.52[**]	.49[**]	.55[**]	.55[**]
Self-disciplinary	.33[**]	.42[**]	.37[**]	.43[**]

Note. IM=Impression Management, SDE=Self-Deceptive Enhancement

[a]n =86.[b]n=75.

[†]$p < .10$, [*]$p<0.05$, [**] $p<0.01$.

conditions. Although not the focus of the present study, the results of the current study agreed that the concern that IM and SDE contaminate with personality scores was valid.

BS inventory's correlations with cognitive abilities. Exploratory question 2 examined the question regarding BS inventory's correlation with cognitive abilities and academic performance. In order to investigate the question, bivariate correlation tests were conducted. The results of the correlation analyses are available in Table 12. Under the motivated condition, the scores of BS inventory were negatively correlated with both ACT (r= -.26, *p*< .01, for dichotomous scores; r= -.23, *p*< .01, for continuous scores), and GPA (r= -.40, -.37, *p*< .01, for dichotomous and continuous scores, respectively). The relationships between BS inventory and ACT/GPA were mitigated under the control condition. The results indicated that BS scores also contain a cognitive component aside from measuring faking, which were in consistent with the findings of Levashina et al. (2009).

Between-subjects analysis of alpha coefficients for OCQ-Foils. The exploratory question 3 examined whether the reliabilities of the OCQ-Foils (i.e., alpha coefficients of the eight foils) would change between the motivated condition and the control condition. To investigate this question, W-tests (Feldt, 1969) were conducted in R using the cocron-package (Diedenhofen, 2013). The results of W-tests are available in Table 13. According to the results, the reliabilities of OCQ-Foils did not change across the lab conditions. The non-significant results may due to that the OCQ-Foils had more items than the BS inventory (i.e., 8 foils, compared to 4 bogus items), which would result in higher reliabilities. The results may also due to that participants did not fake enough on OCQ-Foils. Although OCQ-Foils showed a significant mean difference between the lab conditions, the effect sizes for between-subjects comparison were small (i.e., $\eta^2 = .05$, for dichotomous scores; $\eta^2 = .03$, for continuous scores).

Table 12

Correlation Coefficients Between BS Inventory, GPA and ACT.

Scale	Motivated [a]		Control [b]	
	GPA	ACT	GPA	ACT
BS dichotomous	-.26[*]	-.40[**]	0.11	-0.24[†]
BS continuous	-.23[*]	-.37[**]	0.16	-0.18

Note. [a]$n =61$. [b]$n=58$.

[†]$p < .10$, [*]$p<0.05$, [**] $p<0.01$.

Table 13.

Between-Group Alpha Coefficient Analysis for OCQ-Foils

Scale	Motivated Condition	Control Condition	*W (two-tailed)*
FOILS dichotomous	0.64	0.62	1.06
FOILS continuous	0.76	0.73	1.13

Note. W is a statistical test based on the sampling distribution of coefficient alpha which suggests that the ration (W) is distributed as a central F with dfs (N1-1, N2-1). The degrees of freedom of the above tests were (85, 74).

†$p < .10$, *$p<0.05$, ** $p<0.01$.

Correlation with faking criterion

Hypothesis 4 examines whether the faking scores (i.e., IM, SDE, BS inventory and OCQ-Foils) would demonstrate proxy validity under the motivated condition. The faking criterion was obtained following the method of Paulhus (2002). For each individual, the personality score of each dimension at Time 2 (i.e., the lab condition) was regressed over the corresponding personality score at Time 1 (i.e., the baseline). The faking criterion of each personality dimension was obtained as the residual score of the above regression. In order to test hypothesis 4, bivariate correlation analyses were conducted between the faking scores and the residual scores (i.e., the faking criterion) for participants of the motivated condition. The results of the correlation analyses are available in Tables 14.

According to Table 14, IM and SDE showed significant correlations with the residual scores with most personality dimensions. The correlation coefficients between IM and residual scores were very close to the correlation coefficients between SDE and residual scores. The BS and OCQ-Foils did not show any significant correlation with the faking criterion. The results indicated that IM and SDE showed stronger validity as faking measures than BS and OCQ-Foils. Moreover, the IM and SDE showed equivalent validities as to the proxy validity with the residual scores (i.e., the faking criterion).

Table 14

Correlations between Faking Measures and Residual Scores of Personality, Motivated Condition

Residual Scores	IM	SDE	BS continuous	BS dichotomous	FOILS continuous	FOILS dichotomous
Agreeableness	.38**	.36**	-.25*	-.28*	-0.05	-0.12
Conscientiousness	.50**	.34**	.11	.11	0.15	0.12
Emotional Stability	.51**	.49**	.23‡	.23‡	.26*	0.19
Extroversion	.53**	.37**	.14	.11	0.13	0.05
Openness	.45**	.49**	-.14	-.13	-0.11	-0.09
Achievement striving	.41**	.37**	-.11	-.12	0.07	-0.02
Competitiveness	.48**	.44**	.18	.12	.26*	0.12
Self-Disciplinary	.26*	0.22	.02	-.03	0.02	-0.01
Total score of school conscientiousness	.43**	.39**	.03	-.01	0.1	0.03

Note. n = 62; participants in the motivated condition with mass screening scores were included.

* *p* < .05. ** *p* < .01.

Suppression effects

Hypothesis 5 examines the suppression effects of the faking scores. The suppression effects were tested with multi-step regression analyses following the methods of Bing et al. (2011). Specifically, ACT, as the controller, was entered into the equation in Step 1. The personality score, as a predictor, was added in to the regression equation in Step 2. The faking measure, as a potential suppressor, was added in Step 3. A classical suppression effect requires three conditions to be satisfied (i.e., Paulhus, 2004; Bing et al., 2011). First, the faking score should explain a significant proportion of variances in the dependent variable (i.e., condition 1). Second, the regression coefficient of personality score should increase substantially after adding the faking score (i.e., condition 2). Finally, the faking score should achieve a significant negative correlation with the dependent variable (i.e., condition 3).

Fifty-four sets of regression analyses were conducted, for each faking score (i.e., 6 faking scores) upon the prediction of each personality score (i.e., 9 personality scores). Table 15 summarizes the results of the suppression effects. A detailed description of suppression effect analyses is provided in Table 16. IM and SDE showed several suppression effects upon the prediction of school-specific conscientiousness scores towards GPA. The BS and OCQ-Foils showed no significant suppression effects. Once again, the results indicated that IM and SDE showed higher validity as faking measures than the BS and OCQ-Foils.

Table 15

Summary of Suppression Effects of Faking Measures upon Personalities, Motivated Condition

Suppressors	Agreeable	Conscientious	Emotional Stability	Extroversion	Openness	Achievement Striving	Competitive	Self-disciplinary	Total score of school-conscientious
IM	N	N	N	N	N	**Y**	**Y**	**Y**	**Y**
SDE	N	N	N	N	N	**Y?**	N	**Y**	**Y**
BS dichotomous	N	N	N	N	N	N	N	N	N
BS continuous	N	N	N	N	N	N	N	N	N
FOILS dichotomous	N	N	N	N	N	N	N	N	N
FOILS continuous	N	N	N	N	N	N	N	N	N

Note. Each cell represents a set of suppression effect analysis. "Y" represents that the suppression effect was significant (i.e., all three conditions for the classical suppression effects were satisfied, and R^2 change at Step 3 was significant at the .05 level). "Y?" represents that the suppression effect

was marginally significant (i.e., all conditions satisfied; only the R^2 change at Step 3 was at the .10 level). "N" indicates that at least two conditions were not satisfied.

$n = 75$; all participants in the motivated condition with GPA and ACT scores were included.

Y = suppression effect supported at $p = .05$ level.

Y? = suppression effect supported at $p = .10$ level.

Table 16

Regression Results with IM and SDE Scores as the Suppressors of Personalities Validity, Motivated Condition

Predictors	Step1	Step2	Step3
ACT	.47**	.43**	.45**
Achievement Striving		.30*	.43**
IM			-.36*
R^2	.22**	.29**	.39**
ΔR^2	.22**	.07**	.10**
ACT	.47**	.45**	.48**
Competitiveness		0.14	.34**
IM			-.38**
R^2	.22**	.24**	.31**
ΔR^2	.22**	.02*	.07**
ACT	.47**	.47**	.51**
Self-disciplinary		.29*	.41**
IM			-.34*
R^2	.22**	.30**	.40**
ΔR^2	.22**	.08**	0.10**
ACT	.47**	.45**	.49**
Total score of school specific conscientiousness		.28*	.44**
IM			-.39**
R^2	.22**	.29**	.39**

ΔR^2	$.22^{**}$	$.07^{**}$	$.10^{**}$
ACT	$.47^{**}$	$.43^{**}$	$.40^{**}$
Achievement Striving		$.30^{*}$	$.36^{**}$
SDE			$-.18^{*}$
R^2	$.22^{**}$	$.29^{**}$	$.33^{*}$
ΔR^2	$.22^{**}$	$.07^{**}$	$.04^{\ddagger}$
ACT	$.47^{**}$	$.45^{**}$	$.43^{**}$
Competitiveness		0.14	$.23^{\ddagger}$
SDE			-0.18
R^2	$.22^{**}$	0.24	0.26
ΔR^2	$.22^{**}$	0.02	0.02
ACT	$.47^{**}$	$.47^{**}$	$.45^{**}$
Self-disciplinary		$.29^{*}$	$.38^{*}$
SDE			$-.22^{*}$
R^2	$.22^{**}$	$.30^{**}$	$.34^{*}$
ΔR^2	$.22^{**}$	$.08^{**}$	$.04^{*}$
ACT	$.47^{**}$	$.45^{**}$	$.42^{**}$
Total score of school specific conscientiousness		$.28^{*}$	$.38^{**}$
SDE			$-.23^{*}$
R^2	$.22^{**}$	$.29^{**}$	$.33^{**}$
ΔR^2	$.22^{**}$	$.07^{**}$	$.04^{*}$

Note. Regression coefficients were all standardized. $n=75$. $^{*}p < .05$. $^{**}p < .01$.

Moderation effects

Hypothesis 6 examines the moderation effects of faking measures. The moderation effects were examined through moderated regression analyses, following the approach of Berry et al. (2007). A total of 54 sets of moderated regression analyses were conducted (i.e., 9 personality scores, 6 faking scores). Meanwhile, 54 dummy variables were created, as the production terms between each personality scores and each faking scores. For each moderated regression analysis, the ACT, as a controller, was entered into the regression equation in Step 1. The personality score and the faking score were entered in Step 2. The production term was entered in Step 3. A significant moderation effect requires that the production term at Step 3 to be significant (i.e., a significant R^2 change at Step 3). Table 17 summarizes the results of testing for moderation effects. A detailed description of significant moderation effects is available in Table 18. The IM and SDE showed some significant moderation effects. Specifically, IM moderated the criterion-related validity of conscientiousness and extroversion scores. SDE moderated the validity of achievement striving score and competitiveness score, as well as the school-specific conscientiousness total score. The BS and OCQ-Foils, again, did not show any moderation effects.

Simple slope tests were conducted for significant and marginally significant moderation effects based on the Preach (2010) method (i.e., see Figures 1–5 for the simple slope plots). As hypothesis 6 indicates, the predictive validity of personality scores is expected to be lower when faking score is high, and to be higher when faking score is low. The simple slope tests revealed that all moderation effects came in the expected direction, with one exception (i.e., the interaction between IM and Extroversion). The results indicated that again, the IM and SDE showed higher validity than the BS and OCQ, in regarding to the moderation effects.

79

Table 17

Summary of Moderation Effects of Four Faking Measures, Motivated Condition

Suppressors	Agreeable	Conscientious	Emotional Stability	Extroversion	Openness	Achievement Striving	Competitive	Self-disciplinary	Total score of school specific conscientious
IM	N	**Y?**	N	**Y**	N	N	N	N	N
SDE	N	N	N	N	N	**Y**	**Y?**	N	**Y?**
BS dichotomous	N	N	N	N	N	N	N	N	N
BS continuous	N	N	N	N	N	N	N	N	N
FOILS dichotomous	N	N	N	N	N	N	N	N	N
FOILS continuous	N	N	N	N	N	N	N	N	N

Note. Each cell represents a set of moderated regression analysis. "Y" represents that the moderation effect was significant (i.e., R^2 change at Step 3 was significant at .05 level). "Y?" represents that the moderation effect was marginally significant (i.e., R^2 change at Step 3 was at .10 level). "N" represents that the moderation effect was not significant (i.e., R^2 change at Step 3 was not significant).

$n = 75$; all participants in the motivated condition with GPA and ACT scores were included.

Y = moderation effect supported at $p = .05$ level.

Y? = moderation effect supported at $p = .10$ level.

Table 18

Moderated Regression Results of Faking Measures (with Significant Interactions),

Motivated Condition

Predictors	Step1	Step2	Step3
ACT	.08**	.08**	.07**
Conscientious		0.1	.87‡
IM		-.20*	0.66
Conscientious × IM			-.15‡
R^2	.22**	.27**	.30**
ΔR^2	.22**	.03‡	.03‡
ACT	.08**	.07**	.06**
Achievement-striving		-.19**	1.08**
SDE		-.15‡	.77*
Achievement-striving × SDE			-.17*
R^2	.22**	.33**	.38**
ΔR^2	.22**	.11**	.05*
ACT	.08**	.08**	.08**
Extroversion		-0.07	.72*
IM		-0.11	.79‡
Extroversion × IM			-.16*
R^2	.22**	.26**	.32**
ΔR^2	.22**	0.05	.05*
ACT	.08**	.07**	.08**

Competitiveness		$.14^{‡}$	$.81^{*}$
SDE		-0.15	0.59
Competitiveness $\times$ SDE			$-.13^{‡}$
R^2	$.22^{**}$	$.26^{**}$	$.29^{**}$
ΔR^2	$.22^{**}$	0.04	$.03^{‡}$
ACT	$.08^{**}$	$.07^{**}$	$.06^{**}$
Total score of school specific conscientiousness		$.22^{**}$	$.88^{*}$
SDE		$-.19^{*}$	0.53
Total score of school specific conscientiousness $\times$ SDE			$-.13^{‡}$
R^2	$.22^{**}$	$.33^{**}$	$.36^{**}$
ΔR^2	$.22^{**}$	$.11^{**}$	$.03^{‡}$

Note. Regression coefficients were all un-standardized.

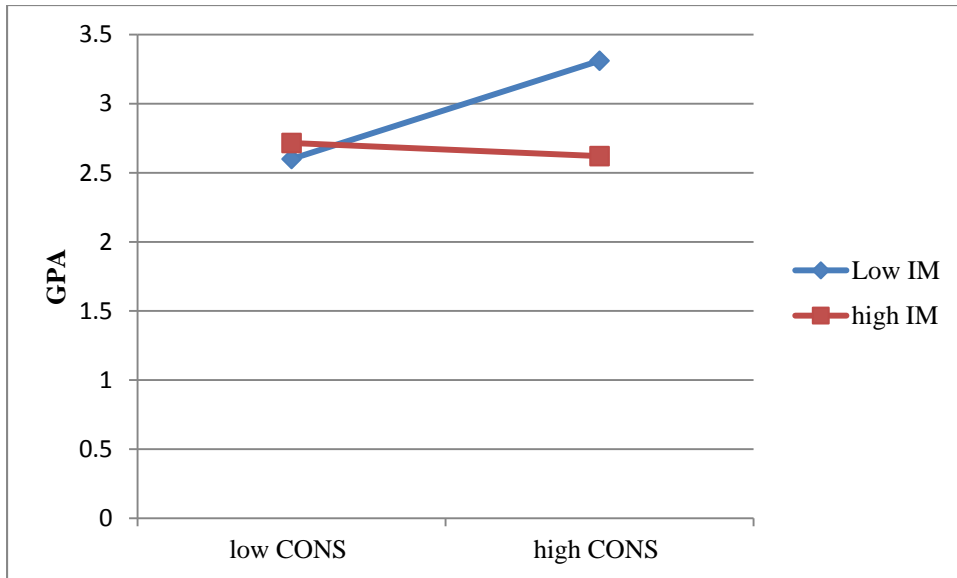$n = 75.$ $^{‡}p < .10.$ $^{**}p < .01.$ $^{*}p < .05.$

Figure 1. Simple slope plot for IM moderates the relationship between IPIP-Conscientiousness and GPA. When IM is one SD below the mean, the simple slope = .22, *p* < .05; when IM is one SD above the mean, simple slope = -.03, *p* > .05.
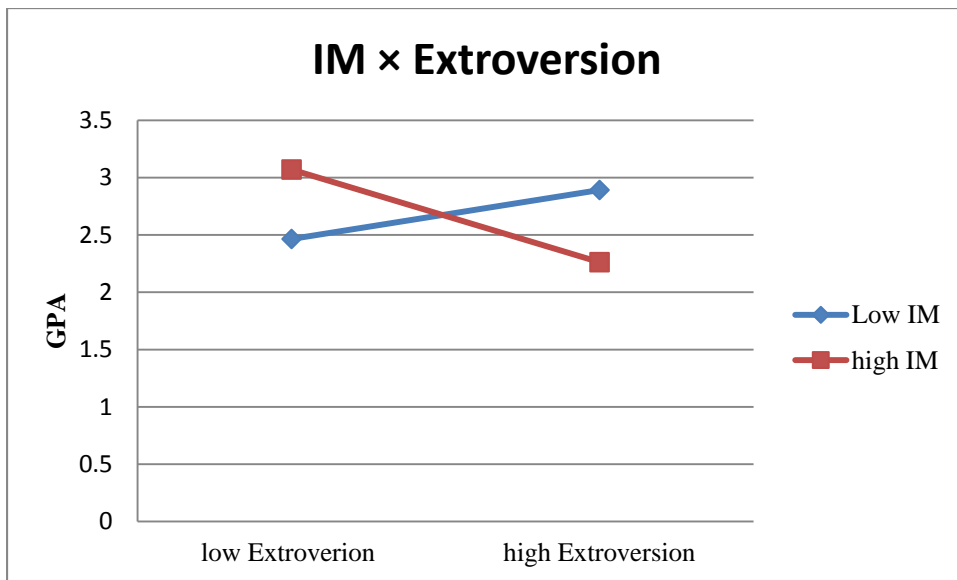

Figure 2. Simple slope plot for IM moderates the relationship between IPIP-Extraversion and GPA. When IM is one SD below the mean, the simple slope = .10, *p* > .05; when IM is one SD

above the mean, simple slope = -.19, $p < .05$. This is the only moderation effect that was not in
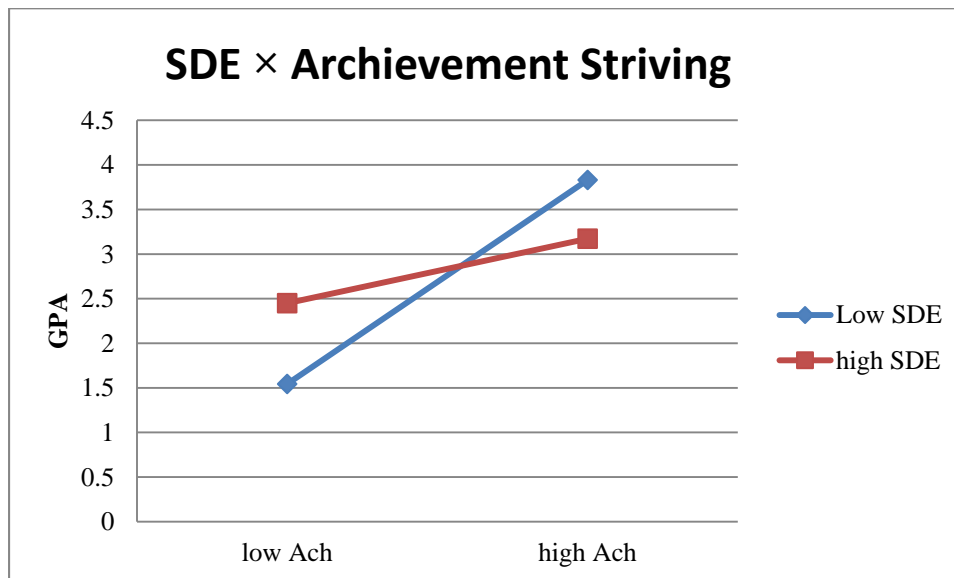
the expected direction.

Figure 3. Simple slope plot for SDE moderates the relationship between Achievement Striving

and GPA. When SDE is one SD below the mean, the simple slope = .44, $p < .01$; when SDE is

one SD above the mean, simple slope = .13, $p > .05$.
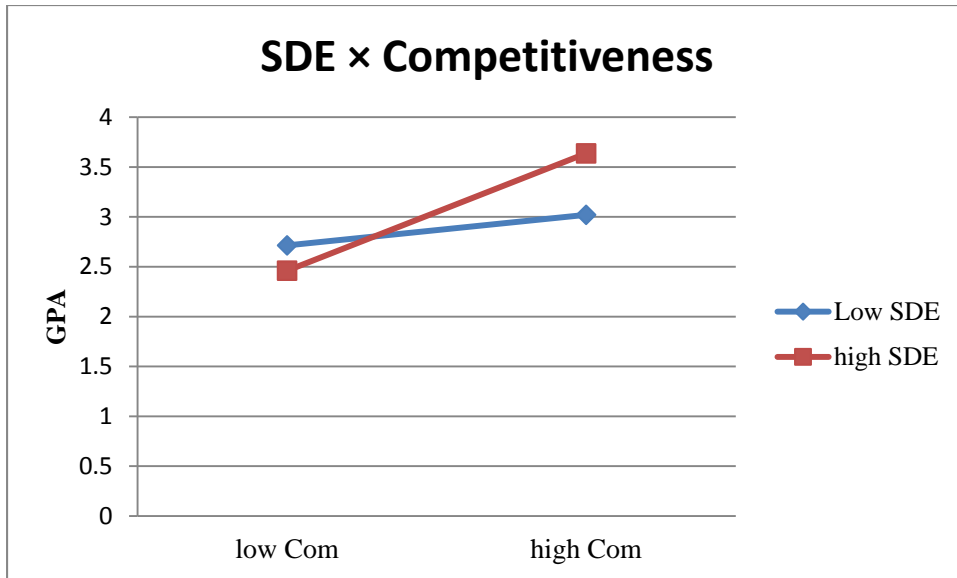
Figure 4. Simple slope plot for SDE moderates the relationship between Competitiveness and GPA. When SDE is one SD below the mean, the simple slope = .30, $p < .05$; when SDE is one SD above the mean, simple slope = .07, $p > .05$.
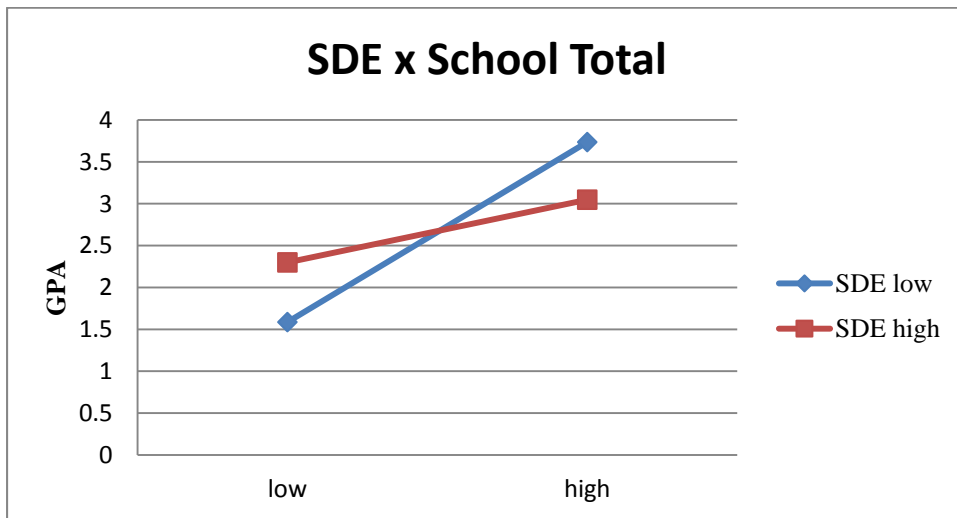

Figure 5. Simple slope plot for SDE moderates the relationship between School Specific Conscientiousness Total Score and GPA. When SDE is one standard deviation below the mean,

simple slope = .36, p<.01. When SDE is one standard deviation above the mean, simple

slope= .12, p>.05.

Comparison of the validities of faking measures

Exploratory Question 4 attempted to compare the validities of the faking measures' (i.e., IM, SDE, BS and OCQ-Foils). In general, the results of hypotheses 4, 5, and 6 indicated that IM and SDE were valid faking measures. The IM and SDE demonstrated their validities in all three criterions: correlations with the residual scores, the suppression effects, and the moderation effects. The results also indicated that BS and OCQ-Foils, on the other hand, lacked validity in all three aspects. BS and OCQ-Foils showed very few significant correlations with the residual scores, and did not show any suppression or moderation effects.

It is difficult to compare the validity between the BS and OCQ-Foils, since both of them showed very little validity in measuring faking. However, it is possible to compare the validities of IM and SDE. To begin with, IM and SDE showed equivalent correlations with the residual scores. Moreover, both IM and SDE showed suppression effects upon predictive validity of school specific contentiousness. The IM moderated all the four school-specific scores, while the SDE moderated three out of the four scores (i.e., with the exception of competitiveness). Finally, both IM and SDE showed moderation effects on some of the personality dimensions. The only difference lies in that the IM moderated the prediction of IPIP personality scores, while SDE moderated the prediction of school-specific conscientiousness scores. However, this difference was not observed with other analyses (i.e., correlation with residual scores, or the suppression effects). Based on the evidences, it can be concluded that the IM and SDE demonstrated equivalent validities in measuring faking.

The Validities of Faking Measures under the Control Condition

Explanatory question 5 and 6 attempted to investigate the validity of the faking measures under the control condition (i.e., a relatively low motivated context). Majority of past studies on

faking were conducted under a low motivated context, and these studies typically found no validity of faking scores (i.e., see McGrath, 2010). The control condition of present study attempted to replicate these results and investigate whether the faking scores would be stil valid under a low motivated context.

The explanatory question 5 explored the correlation between faking scores and residual scores of personalities (i.e., the faking criterion) under the control condition. The residual scores for the control condition were obtained in the same way as for the motivated condition. The residual scores were obtained by regressing the equated personality scores at Time 2 over the Time 1 scores. Bivariate correlation analyses were conducted to obtain the correlations between the faking scores and the residual scores, and the results are available in Table 19. The IM and SDE showed significant correlations with the residual scores. However, the number of significant correlations was reduced. The sizes of correlation coefficients were also reduced. Similar to the motivated condition, the BS inventory and OCQ-Foils showed no correlation with the faking criterion. In general, the results indicated that consistent with the speculation, the faking scores correlated less with the residual scores under the control condition.

The explanatory question 6 investigated whether the faking scores would show suppression and moderation effects under the control condition. Suppressed regression and moderated regression analyses were conducted for the control condition following the same methods as for the motivated condition. A summary of the suppression effects is available in Table 20, and see Table 21 for the moderation effects. None suppression effects were observed under the control condition. Only one of fifty-four sets of moderated regression analyses was marginally significant (i.e., dichotomous BS upon the prediction of openness; see Table 22 for the details). The simple slope test following the methods of Preacher et al. (2006) revealed that

the simple slopes were only significant outside the range of the BS scores. As a result, the simple slope analysis was conducted by the split-files in SPSS. Firstly, participants in the control condition whose dichotomous BS scores were below the mean were selected (i.e., N=48). A simple regression analysis of Openness towards GPA was conducted for these participants, controlling for participants' ACT. When participants' BS scores were below the mean, the prediction of Openness towards GPA was marginally significant above and beyond students' ACT, Beta=.20, p<.10. Next, participants in the control condition whose dichotomous BS scores were above the mean were selected (i.e., N= 10). Another simple regression analysis of Openness towards GPA was conducted for these participants, controlling for participants' ACT. When participants BS scores were above the mean, the prediction of Openness towards GPA was no longer significant above and beyond ACT, Beta=-.24, p>.05. The results revealed that the direction of the moderation effects was still in the predicted direction, i.e., the prediction of Openness was higher when BS scores were low, and was lower when BS scores were high.

Table 19

Correlations Between Faking Measures and Residual Personality Scores Under Control

Condition

Residual Scores	IM	SDE	BS continuous	BS dichotomous	FOILS continuous	FOILS dichotomous
Agreeableness	.34**	.30*	.10	.04	.27*	.15
Conscientiousness	.36**	.29*	-.02	.04	.00	-.07
Emotional Stability	.43**	.53**	-.21	-.19	.04	.03
Extroversion	.07	.25	-.17	-.19	.05	.05
Openness	.02	.17	.02	.01	.16	.17
Achievement striving	.20	.45**	-.02	.06	-.11	-.13
Competitiveness	.42**	.50**	-.15	-.10	-.04	-.01
Self-Disciplinary	.24	.40**	.12	.13	-.01	.00
Total score of-school conscientiousness	.25	.49**	.01	.05	-.05	-.05

Note. *n*=57; participants in the control condition with mass screening scores were included.

* $p < .05.$ ** $p < .01.$

Table 20

Summary of Suppression Effects of Faking Measures, Control Condition

Suppressors	Agreeable	Conscientious	Emotional Stability	Extrovert	Openness	Achievement Striving	Competitive	Self-disciplinary	Total score of school conscientious
IM	N	N	N	N	N	N	N	N	N
SDE	N	N	N	N	N	N	N	N	N
BS dichotomous	N	N	N	N	N	N	N	N	N
BS continuous	N	N	N	N	N	N	N	N	N
FOILS dichotomous	N	N	N	N	N	N	N	N	N
FOILS continuous	N	N	N	N	N	N	N	N	N

Note. Each cell represents a set of moderated regression analysis. "Y" represents that the suppression effect was significant (i.e., all three conditions for the classical suppression effects were satisfied, and R^2 change at Step 3 was significant at the .05 level). "Y?" represents that the suppression effect was marginally significant (i.e., all conditions satisfied; only the R^2 change at Step 3 was at the .10 level). "N" indicates that at least two conditions were not satisfied.

$n = 60$; all participants in the control condition with GPA and ACT scores were included.

Y = moderation effect supported at $p = .05$ level.

Y? = moderation effect supported at $p = .10$ level.

Table 21

Summary of Moderation Effects of Faking Measures, Control Condition

Suppressors	Agreeable	Conscientious	Emotional Stability	Extrovert	Openness	Achievement Striving	Competitive	Self-disciplinary	Total score of school conscientious
IM	N	N	N	N	N	N	N	N	N
SDE	N	N	N	N	N	N	N	N	N
BS dichotomous	N	N	N	N	**Y**	N	N	N	N
BS continuous	N	N	N	N	N	N	N	N	N
FOILS dichotomous	N	N	N	N	N	N	N	N	N
FOILS continuous	N	N	N	N	N	N	N	N	N

Note. Each cell represents a set of moderated regression analysis. "Y" represents that the moderation effect was significant (i.e., R^2 change at Step 3 was significant at .05 level). "Y?" represents that the moderation effect was marginally significant (i.e., R^2 change at Step 3 was at .10 level). "N" represents that the moderation effect was not significant (i.e., R^2 change at Step 3 was not significant).

$n = 60$; all participants in the control condition with GPA and ACT scores were included.

Y = moderation effect supported at $p = .05$ level.

Y? = moderation effect supported at $p = .10$ level.

Table 22

Moderated Regression Results of Faking Measures, Control Condition

Predictors	**Step1**	**Step2**	**Step3**
ACT	$.09^{**}$	$.09^{**}$	$.08^{**}$
Openness		$.11$	$1.80^{‡}$
BS dichotomous		$1.46^{‡}$	9.33^{*}
Openness × BS dichotomous			$-.59^{‡}$
R^2	$.20^{**}$	$.27^{**}$	$.31^{**}$
ΔR^2	$.20^{**}$	$.07^{‡}$	$.04^{‡}$

Note. Regression coefficients were all un-standardized.

$n = 63$. $^{‡}p < .10$. $^{**}p < .01$. $^{*}p < .05$.

Discussion

The purpose of the present study is to examine the validity of faking measures under lab environment. The validities of four faking measures (i.e., IM, SDE, BS and OCQ-Foils) were tested under both the simulated selection context (i.e., the motivated condition), and the research context (i.e., the control condition). Under motivated condition, IM and SDE correlated substantially correlations with the residual scores of personalities (i.e., the faking criterion). IM and SDE also showed suppression and moderation effects upon the criterion-related validity of personality scores. While BS and OCQ neither correlated with the residual scores, nor showed suppression or moderation effects. Under the control condition, IM and SDE continued to correlate with the residual scores. The BS and OCQ did not correlate with the residual scores and did not show suppression and moderation effects. None suppression effects was observed under the control condition. Only one set of marginally significant moderation effect was observed. In summary, the results indicated that IM and SDE were more valid than BS and OCQ-Foils in measuring faking. The fact that the traditional faking measures (i.e., IM and SDE) showed more validities than the faking measures of new approach (i.e., BS inventory and OCQ) was quite consistent throughout the analyses and has been observed repeatedly across different analyses.

Despite of the consistency of the overall results, there are still some important questions to ask about the results. For example, why would IM and SDE show correlation with the faking criterion? Why the BS and OCQ-Foils were less valid than IM and SDE? How to explain the suppression and moderation effects? The discussion section is aimed to address these questions. To begin with, the faking scores' validities will be discussed, in all three aspects (i.e., correlations with residual scores, suppression effects, moderation effects), and for both motivated

condition and control condition.   Moreover, the limitations of the present study will be addressed, and the future implications will be provided.

Discussion of faking scores' correlations with the faking criterion

For a long time, the within-subjects score change has been viewed as the "Golden criterion" for test faking (Griffith et al., 2011; Peterson & Griffith, 2008). According to Peterson and Griffith (2008), a significant correlation between faking measures and the faking criterion would indicate the "proxy validity" of the faking measure, which represents the similarity between a measurement and the construct it attempts to capture. In hypothesis 4, I examined whether these four faking measures would show such proxy validity under a simulated selection context.   In my examination, only the IM and SDE showed significant correlations with the faking criterion, while measures of the new approach (i.e., BS and OCQ-Foils) did not show proxy validity. The results were opposite to the findings of Peterson et al. study (2011). Under an application context, Peterson et al. (2011) found that SDR scales (i.e., the Marlowe-Crowne scale in Peterson's study) showed no significant correlations with the residual scores. This is the only study that examined the correlation between SDR scales and the faking criterion under a high motivated context.

Why would IM and SDE show proxy validity with the faking criterion? To begin with, it is possible that correlations among IM, SDE, and personality scores were due to the fact that the three variables captured the same construct—the inflated test motivation. Previous studies have revealed that individuals tend to score higher on IM and SDE under high-motivated test context (Smith & Ellingson, 2002; Bing et al., 2011; Holden & Evory, 2005; Dodaj, 2012; Paulhus & Rein, 1991, Pauls & Crost, 2004). Research also indicated that applicants can score much high on personality tests than non-applicants (i.e., Ones et al., 1996). In the present study, the residual

scores represented the inflated test motivation across baseline assessment (i.e., Time 1) and lab

condition (i.e., Time 2). Since the scores of IM and SDE were impacted by the test motivation, it

is possible that the scores of IM and SDE can correlate with the residual scores due to the shared

impact of test motivation.

Moreover, the correlation between IM, SDE and personality scores can also be explained

by the common method variance. As mentioned in the Methods part, the items of IM and SDE

are very similar to personality items. Both of them composed of a statement of one's typical

behavior and ask for the level of agreement with the statement.   At Time 2, the items of IM and

SDE were mixed with personality items and were administrated to participants in a survey using

the same rating scale.   As a result, there can be considerable common method variance between

Time 2 personality scores and IM/SDE scores, which can carry over to the residual scores. The

common method variance also helps to explain why IM and SDE were correlated with

personality residual scores under the control condition. It is possible that the correlations

between IM, SDE, and residual scores of the control condition were accounted for by the

common method variances, too, for the reason that less faking occurred in the control condition.

Why the BS and OCQ-Foils did not show significant correlation with the residual scores?

In general, my tentative explanation was that participants did not fake enough on BS and OCQ-

Foils. To begin with, both BS and OCQ-Foils had a very low mean score, even under the

motivated condition. Under the motivated condition, the mean for dichotomous BS was

around .40 for dichotomous scoring (i.e., on average participants endorsed 0.4 out of 4 bogus

items per person). The mean score for continuous BS was only slightly higher than 1 (i.e.,

indicating only slightly higher than "never" in the frequency of obtaining the bogus items). The

situation was similar with OCQ-Foils (i.e., see Table 6 for more details). These numbers

indicated that participants of the motivated condition did not fake enough on those two scales. Moreover, although participants scored significantly higher on BS and OCQ-Foils at the motivated condition compared with the control condition, the effect sizes were relatively small (i.e.,). In summary, less-than-adequate faking on BS and OCQ-Foils may have rendered BS and OCQ-Foils captured less constructs regarding applicants faking and test motivation.

Discussion of the suppression effects

Researchers have long been assumed that faking scores should show suppression effects upon the prediction of personality scores (e.g., Ones, et al., 1996; Ganster, 1986). However, there has not been a single study observed such a suppression effect of SDR scales (i.e., including IM and SDE) upon the predictive validity of personality scores. Nevertheless, the present study observed a suppression effect with the SDR scales. In motivated condition, IM and SDE showed several significant suppression effects upon the predictive validity of the school-specific conscientiousness scores. The BS and OCQ-Foils did not show any suppression effects. No suppression effects were observed under the control condition.

There are several potential explanations why the present study observed suppression effects with the SDR scales. To begin with, majority of past studies on faking and SDR were not done properly (e.g., McGrath, 2010; Bing et al., 2011). They were either conducted under low motivated context, or tested the suppression effects with the semipartial correlation approach. Therefore, suppression effects were not observed in the past studies can be partly due to these factors. Moreover, the observation of suppression effects requires a moderate to large prediction of personality scores towards the performance criterion. In the present study, the suppression effects were only observed with school-specific conscientiousness scores, but not with the IPIP scores. In fact, under both lab conditions, none of the IPIP scores achieved a significant zero-

order correlation with GPA.   In contrast, under the motivated condition, two out of three school-specific scores (i.e., achievement striving, competitiveness) showed significant zero order correlation with GPA. Since the school-specific scores are more relevant to the criterion, they should demonstrate higher predictive ability towards the GPA. The above results also reflected the theory of frame of references (Bing et al., 2004). The results also highlighted the importance of using predictive personality sores in the faking research.

Previous researches have observed the suppression effects of BS inventory (Anderson et al., 1984) and OCQ-Foils (Bing et al., 2011). However, these results were not replicated in the present study. These results may be explained, again, by that participants did not fake enough on these scales. The BS and OCQ-Foils were unable to suppress the criterion-related validity of personality scores due to less than enough faking construct captured.   In the future, more studies on BS inventory and OCQ are needed to provide further evidences regarding their suppression effects.

Discussion of the moderation effects

The moderation effects of faking scores have been overlooked in the faking literature. Up to the date, only four published studies examined, and found significant moderation effects of faking scores (i.e., Hough, 1991; Berry et al., 2007; Lanyon, 2014; White et al., 2001). Most of the above studies tested the moderation effects under a low motivated context (e.g., Hough, 1991; Berry et al., 2007). The present study offered to examine the moderation effects of faking scores under both high-and-low motivated contexts.

In the motivated condition, IM and SDE showed significant moderation effects upon the criterion-related validities of personality scores. The moderation effects of SDE occurred with the school-specific conscientiousness scores. The moderation effects of IM occurred with

91

conscientiousness and extroversion scores. The simple slope tests revealed that the moderation effects were in the expected direction, such that the predictions of personality scores were higher when IM/SDE scores were low, and were lower when IM/SDE scores were high. The BS and OCQ-Foils, however, failed to show any moderation effects. Under the control condition, only BS showed a marginal significant moderation effects towards the prediction of openness. A split-file regression analysis revealed that the direction of the moderation effects was still under the expected direction, that is, the prediction of openness was higher when BS was low, and was lower when BS was high.

There has been lacking explanations regarding when, how, and why different SDR scales should moderate the predictive validity of personality dimensions. One potential theory in explaining the moderation effects of faking measures is the Big-Two personality theory (Pauhus & John, 1998; Gebauer, Paulhus, & Neberich, 2012). The Big-Two theory divide personality traits into two fundamental dimensions—a communal dimension which represents warmth, relatedness, morality, as well as a motivation for assimilation, and an agentic dimension, which presents competence, uniqueness, ambition, and a differential motivation (Gebauer et al., 2012). In regarding to the Big-Five traits, the agreeableness and conscientiousness are communal traits, while emotional stability, extroversion, and openness are agentic traits (see Gebauer et al., 2012; Paulhus, 2002). The same rule also applies to social desirability. As mentioned in the literature review part of this thesis, IM represents a communal bias, while the SDE represents an agentic bias.

The Big-Two theory has been applied to explain the occurrence of moderation effects with IM and SDE. For example, Berry et al. (2007) used Paulhus' Big-Two theory to explain the moderation effect of his study. Berry et al. found that SDE moderated the prediction of

92

extroversion towards leadership performance. Berry believed that the reason why SDE and extroversion showed a significant interactive effect lies in two aspects. Firstly, extroversion is a strong predictor of leadership performance. Secondly, since both extroversion and SDE are associated with agentic trait/bias, it makes sense that they interact with each other.

However, Berry's arguments would not apply in the present study. To begin with, the moderation effects were not only observed with school-specific conscientiousness, which were strong predictors of college students' GPA, but also with IPIP Big-Five dimensions. Moreover, the moderation effects found were not in congruent with Paulhus' theory.   As communal traits, the school specific conscientiousness scores interacted with SDE, which is associated with agentic bias. The results that IM moderated the prediction of emotional stability also violated Paulhus' theory, in which emotional stability is an agentic trait, and IM to be a communal bias. More studies are needed to further examine the boundary conditions of the occurrence of moderation effects.

Discussion over faking scores' validities under the control condition

Results regarding the control condition found less evidence regarding the validity of faking scores under the low motivated context. Under the control condition, IM and SDE scales still correlated, although to a less extent, with the faking criterion. Moreover, the faking scores appear less able impact the predictive validity of the personality scores under the control condition, compared to the motivated condition. Less suppression and moderation effects were observed in the control condition. There was missing a consistent pattern among the observed moderation effects, either. Considering the potential impacts of inflated familywise Alpha in the moderated regression analyses, it was possible that the only significant moderation effect observed under the control condition was only occurred by chance.

93

In general, results of the control condition revealed that the validities of faking scores were largely mitigated, compared to the motivated condition.   The mitigation of the validities of faking scores can be accounted for by the fact that less faking occurred in the control condition. These results further supported the success of experimental manipulation, that is, faking only occurred in the motivated condition, but not in the control condition. Although IM and SDE still captured the inflated test motivation (i.e., the IM and SDE scores correlated with residual scores), they were no longer able to impact the prediction of personalities. The results indicated that most of the score inflation from Time 1 to Time 2 at the control condition was not due to intentional faking.

Limitations

Familywise inflated Alpha for suppressed and moderated regression analyses

The present study examined the effects of four faking measures upon several personality dimensions, many analyses were conducted for testing the suppression and moderation effects. As a result, there raised the concern for family wise inflated alpha, that is, the type I error will increase as the number of family-wise analysis increases. However, under a closer examination, the issue of family-wise alpha inflation may not be as serious as it would.

The family-wise inflated alpha indicates that when the number of family wise analysis increases, a certain proportion of the test would get significant by chance. One assumption that can be made from this point is that, if significant results were only occurred by chance, they should occur randomly and with no observable pattern.   However, this is not the case in the present study.   In fact, the results of the present study indicated a clear pattern that occurred repeated across different type of analyses. To begin with, the significant suppression and moderation effects only occurred with the traditional faking measures—the IM and SDE. None of the

significant test occurred with BS inventory and OCQ-Foils. This pattern is incongruence with the results of hypothesis 4, in which IM and SDE correlated significantly with the faking criterion, while BS inventory and OCQ-Foils did not. Given the pattern that has occurred repeatedly, it is difficult to convince that the significant tests only occurred by chance.   In addition, the suppression effects for IM and SDE only occurred on school-specific conscientiousness dimensions. While for IM, the moderation effects occurred on IPIP dimensions, and on school-specific dimensions for SDE. The fact that significant tests only occurred on certain personality dimensions also indicated that the occurrence of test significance was not merely random, or by chance. All in all, although there is an issue of inflated family-wise alpha, the above evidences have indicated that the results were not likely to have occurred by chance, and in general, were credible.

The deception technique

In general, the lab manipulation has been successful in the present study. Following Ellingson et al. (2012) approach, participants in the motivated condition were told a cover story that they can get a potential job opportunity while completing the questionnaires.   As a result, participants increased their personality scores successfully in the motivated condition, and they also scored higher on all faking measures than participants of the control condition. Given the difficulty of conducting faking research with real job applicants, the deception technique is a good approach which makes it possible to investigate faking under lab conditions.

However, one potential limitation of the deception technique lies in the psychological realism. In the present study, the IM and SDE demonstrated enough validity as faking measures, yet the BS inventory and OCQ-Foils did not—which was due to the fact that participants did not fake on those scales. In a retrospective analysis, there raised the concern regarding why the new

measures, which were objective and has been previously validated(Anderson et al., 1984; Pannone, 1984; Bing et al., 2011), did not work in the present study.

There are a few explanations regarding why participants of the present study did not fake on the measures of the new approach (i.e., BS, OCQ). First, research has indicated that in general, participants are less likely to fake on verifiable faking measures, than the verifiable measures (Morgeson, Delaney-Klinger, Mayfield, Ferrara & Campion, 2004). This is due to the fact that ability statements are less verifiable than task statements. Because the participants were currently enrolled college students and they participated in a non-confidential research study on campus, the concern of verifiability may be why they faked less on the BS inventory and OCQ, but increased their scores more on personalities and SDR scales (i.e., IM, SDE).

Second, the test motivation was not extremely high in the present sample. Only one third of the participants of the motivated condition indicated their interest in the job offered. Under such a context, most of participants may just want to keep honest, rather than take the risk to fake on those objective measures. Third, whether to fake on the objective faking measures can also depend on participants' interpretation of the job criterion. Since the cover story told did not emphasizing on the necessity of specific skills and a variety of knowledge, even highly motivated participants may view the BS inventory and OCQ as something irrelevant to the job content, and thus failed to see the necessity in faking good on them. In conclusion, the relatively high cost of faking on the BS inventory and OCQ (i.e., being verified) as well as the relatively low benefit of faking on those scales (i.e., not very high test motivation) has led participants to neglect the affability of those items and only fake on personalities and "unlikely traits", rather than task or knowledge related inventories.

In addition to the above arguments, one recent study examined the correlation between OCQ and the faking criterion found a similar as the current study (Feeney & Goffin, 2015). The study of Feeney and Goffin (2015) examined the proxy validity of OCQ-90 (Paulhus et al., 2013) against the residual scores of extroversion obtained under a simulated application context, and an honest context. Under the simulated application context, participants were instructed to response as if they were applying for a sales position in sports clothing store (i.e., a yoga store for females). There were also monetary incentives for scoring high on the scales. In their results, the OCQ-Foils only showed a weak correlation with faking criterion of extroversion (i.e., r=.17, p<.01), and the result was significant due to their large sample size (i.e., n=208). This evidence indicated that indeed, it is quite difficult to observe a correlation between the new faking measures and the faking criterion. The measures of the new approach (i.e., BS and OCQ) may only work under certain conditions, such as an obvious connection between the item content and the job criterion, and a high test motivation. These conditions, although with a few exceptions (e.g. Bing et al., 2011), are difficult to simulate under the lab context.

Future Implications

Measuring applicants faking: traditional method, or new approach?

In my perspective, no absolute answer could be given at this time. On one hand, majority of previous studies on IM and SDE failed to support their validities as faking measures (e.g., Li & Bagger, 2006). The present study contributed to the faking literature by adding a piece of evidence in which IM and SDE scales can measure faking, and impact the validity of personality scores. Along with Berry (2007), Lanyon (2014), and White (2001), the present study supported

97

the moderation effects of IM and SDE upon the prediction of personality scores. This is also the first study in the literature ever found suppression effects of IM and SDE scales. Although more evidences are need to further support the validities of IM and SDE in measuring faking, the present study indicated that the traditional measures are not necessarily invalid in measuring faking.

On the other hand, previous studies have found the suppression and moderation effects of BS inventory and OCQ-Foils (e.g., Anderson et al., 1984; Pannone et al., 1984; Bing et al., 2011). Although the present study did not observe the validities of these measures due to a lack of faking on these scales, the single study cannot deny the potential application of the new approach in practices. Although it is premature to make any conclusion at this point, the results of the present study may indicate potential boundary conditions in which the objective faking measures would work. Tentatively, I propose that the objective faking measures are not suitable for use in general entry-level job recruitment, but are rather more suited for skill-based and advanced job types. Moreover, these measures are not likely to work unless the test motivations are really high. Reexamining the literature, I noticed that previous studies reported significant moderation or suppression effects of the BS inventory were based on real job applicants for professional or skill-based jobs and not non-skill based entry-level jobs. With a sample of job applicants in 13 occupational classes (i.e., ground keeper, clerks, secretary, custodian, typist, etc.), Anderson (1984) found a significant suppression effect of the BS inventory on the criterion-related validity of a self-reported bio-date measure. Similarly, using a sample of job applicants in electrician positions, Pannone (1984) found a significant moderation effect of the BS inventory upon the prediction of a bio-date measure. As an example referring to OCQ, Bing et al. (2011) used a college student sample under simulated college application condition and

found a significant suppression effect of OCQ-foils upon the criterion-related validity of personality and students' GPA. Given the evidence, it seems that in order for the BS inventory and OCQ to work, participants have to be highly motivated to get the job offer, while simultaneously there must be situational cues strong enough to link the job position with task-based inventories.

Bogus statement: measuring faking or cognitive ability?

The present study observed a significant correlation between the scores on bogus items (i.e., scores on BS inventory), and the cognitive ability. Under the motivated condition, both the BS scores were negatively correlated with GPA and ACT scores. The negative correlations between dichotomously scored BS inventory and ACT sustained even under the control condition. This finding was in congruence with Levashina et al. (2004), which reported a negative correlation between BS scores and cognitive abilities. In practices, this relationship with cognitive abilities may be a concern, for two aspects. To begin with, the relationship between BS inventory and cognitive abilities cast doubt on the construct validity of BS inventory, which as a measure for faking, are not supposed to correlate with any constructs aside from faking and test motivation. Moreover, this contamination may cause adverse impact as well. In practices, it is necessary to control for the cognitive components of BS inventories and apply BS inventory in a way that would reduce its impact on the cognitive abilities.

In conclusion, the present study indicated that when given the test motivation, faking would occur, and it would matter. An examination of the validity of four faking measures indicated that traditional faking measures (i.e., IM an SDE) worked better than measures of the false item approach (i.e., BS inventory and OCQ-Foils). The IM and SDE showed significant correlations with the faking criterion, and also impacted the predictions of the personality scores

through both suppression and moderation effects.   The BS and OCQ-Foils did not show validity

in the above criterion, which may due to the limitation of experiment manipulation. The study

urged a re-evaluation of SDR scales' validities in measuring faking, while also calls for a

reconsideration of measuring faking with the false item approach.

References

Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, *69*, 574-580.

Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, *18*, 1-16.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of applied psychology*, *81*, 261-272.

Berry, C. M., Page, R. C., & Sackett, P. R. (2007). Effects of self-deceptive enhancement on personality-job performance relationships. *International Journal of Selection and Assessment, 15*, 94-109.

Bing, M. N., Kluemper, D., Davison, H. H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes, 116*, 148-162.

Bing, M. N., LeBreton, J. M., Davison, H. K., Migetz, D. Z., & James, L. R. (2007). Integrating Implicit and Explicit Social Cognitions for Enhanced Personality Assessment A General Framework for Choosing Measurement and Statistical Methods. *Organizational Research Methods*, *10*, 346-389.

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

Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (BIMI): Measuring agentic and communal forms of impression management. *Journal of personality assessment, 96,* 523-531.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6,* 199-214.

Brumback, G. B. (1969). A note on criterion contamination in the validation of biographical data. *Educational and Psychological Measurement*, 29, 439-443.

Byrne, D. (1961). The repression-sensitization scale: Rationale, reliability, and validity. *Journal of Personality, 29*, 334-349.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

Damarin, F., & Messick S. (1965).Response styles as personality variables: A theoretical integration (ETS RB 65-10). Princeton, NJ: Educational Testing Service.

Davis, C. G., Thake, J., & Weekes, J. R. (2012). Impression managers: Nice guys or serious criminals? *Journal of Research in Personality*, *46*, 26-31.

Diedenhofen, B. (2013). cocron: Statistical comparisons of two or more alpha coefficients (Version 1.0-0). *Available from http://r.birkdiedenhofen.de/pckg/cocron/*

Dodaj, A. (2012). Social desirability and self-reports: Testing a content and response-style model of socially desirable responding. *Europe's Journal of Psychology*, 8, 651–666.

Dudek, F. J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin*, *86*, 335-337.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437-448.

Compare ACT & SAT Scores. (2008, June). Retrieved March 25, 2015, from http://www.act.org/solutions/college-career-readiness/compare-act-sat/

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, *24*, 349-354.

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

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

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait are endorsed. *The Journal of Applied Psychology, 37*, 90-93.

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

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: insights into response distortion. *The Journal of applied psychology*, *92*, 386-395.

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin, 43,* 385-440.

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

Fan, J., Wong, C. C., Carroll, S. A., & Lopez, F. J. (2008). An empirical investigation of the influence of social desirability on the factor structure of the Chinese 16PF. *Personality and Individual Differences, 45*, 790-795.

Fan, J., Zhao, G., Li, Y., Wang, M., & Meng, H. (2014, May). Personality scores across selection and development contexts: Evidence for faking. Paper presented as part of the symposium by Fan, J. & Page, R. C. (Chair), *Does faking occur and matter? It's about when, not whether* at the 28th annual conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Feeney, J. R., & Goffin, R. D. (2015). The Overclaiming Questionnaire: A good way to measure faking?. *Personality and Individual Differences*, *82*, 248-252.

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychomelrika*, *34*, 363-373.

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika*, *45*, 99-105.

Fleming, P., & Zizzo, D. J. (2011). Social desirability, approval and public good contribution. *Personality and Individual Differences*, *51,* 258-262.

Fisher, R.A. (1915). "Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population". *Biometrika (Biometrika Trust). 10*, 507–521

Ganster, D.C., Hennessey, H., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal, 26*, 321-331.

Goldstein, L L. (1971). The application blank: How honest are the responses? *Journal of Applied Psychology*, 55, 491-492.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96. doi:10.1016/j.jrp.2005.08.007

Gough, H. G. (1952). On making a good impression. *The Journal of Educational Research*, 46, 33-42.

Gough, H. G. (1987). *CPI Administrator's Guide*. Palo Alto, Calif: Consulting Psy-chologists Inventory Press Inc.

Gough, H. G., & Bradley, P. (1996). *CPI Manual*. Palo Alto.

Griffith, R. L., Malm, T., Yoshita, Y., English, A., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive biases, and individual differences. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 151-178). Greenwich, CT: Information Age Publishing.

Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 308-311.

Griffith, R. L., & Robie, C. (2012). Personality testing and the "F word": Revisiting seven questions about faking. In N. Christiansen & R. Tett (Eds.), *Handbook of Personality at Work* (pp. 253-280). New York: Routledge.

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

Gulliksen, H. (1950). *Theory of Mental Test*. New York: Wiley.

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37, 147-169.

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

Holden, R. R. (2008). Underestimating the effects of faking on the validity of self-report personality scales. *Personality and Individual Differences*, *44*, 311-321.

Holden, R. R., & Evoy, R. A. (2005). Personality inventory faking: A four-dimensional simulation of dissimulation. *Personality and individual differences*, *39*, 1307-1318.

Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences*, *49*, 446-450.

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

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

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

Lanyon, R. I., Goodstein, L. D., & Wershba, R. (2014). 'Good impression' as a moderator in employment-related assessment. *International Journal of Selection and Assessment, 22*, 52-61

Levashina, J., Morgeson, F. P., & Campion, M. A. (2009). They don't do it often, but they do it well: Exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Assessment*, *17*, 271-281.

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*, 131-141.

Klein, S. P., and Owens, W. A. (1965). Faking of a scored life history blank as a function of criterion objectivity. *Journal of Applied Psychology*, 49, 452-454.

Meehl, P., & Hathaway, S. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *The Journal of applied psychology*, *30*, 525-564.

Mueller-Hanson, R., Heggestad, E. D., & Thornton III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.

Nathanson, C., & Paulhus, D. L. (2005, June). *Accuracy and bias in lay knowledge* . Poster presented at the meeting of the Canadian Psychological Association, Calgary .

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.

Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening

applicants. *Personnel Psychology*, *37*, 507-514.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of*

Personality and Social Psychology, 46, 598-609.

Paulhus, D. L. (1988). *Manual for the Balanced Inventory of Desirable Responding: Version 6.*

Unpublished manual, University of British Columbia, Vancouver, British Columbia,

Canada.

Paulhus, D. L. (1998). *Paulhus deception scales (PDS): The balanced inventory of desirable*

responding–7 user's manual. North Tonawanda, NY: Multi-Health System.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I.

Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and*

educational measurement (pp. 49-69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (2006). A Comprehensive Inventory of Desirable Responding (CIDR). In *Poster*

presented at the meeting of the Association for Research in Personality. New Orleans.

Paulhus, D. L. (2011). Overclaiming on personality questionnaires. In Ziegler, M., MacCann,

C., & Roberts, R. (Eds.), *New perspectives on faking in personality assessment* (pp. 151-

164). Oxford University Press.

Paulhus, D. L., & Bruce, N. (1990, June). Validation of the OCQ: A preliminary study. *Paper*

presented at the annual convention of the Canadian Psychological Association, Ottawa,

Ontario, Canada.

Paulhus, D. L., & Dubois, P. J. (2014). Application of the overclaiming technique to scholastic

assessment. *Educational and Psychological Measurement*, *74*, 975-990.

Paulhus, D.L., & Harms, P. D. (2004).Measuring cognitive ability with the over-claiming technique. *Intelligence, 32*, 297-314.

Paulhus, D.L., Harms, P., Bruce, M.N., & Lysy, D. (2001). The Over-Claiming Questionnaire as a proxy measure of cognitive ability and self-enhancement: A meta-analysis. Unpublished manuscript, University of British Columbia, Vancouver, British Columbia, Canada.

Paulhus, D. L., Harm, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 681-693.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1024-1060.

Paulhus, D. L., & Notareschi, R. F. (1993) .Varieties of faking manipulations. Unpublished data, University of British Columbia.

Paulhus , D. L. , & Williams , K. M. (2002). The Dark Triad of personality: narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556-563.

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences*, *37*, 1137-1151.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, *35*, 263-275.

Peterson, M.H., Griffith, R.L., & Converse, P.D. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single and multiple predictor selection. *Journal of Business and Psychology 24*, 373-386.

Peterson, M. H., Griffith, R. L., Isascson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance, 24*, 270-290.

Randall , D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, 10, 805-817.

Raubenheimer , A. S. (1925). An experimental study of some behavioral traits of the potentially delinquent boy. *Psychological Monographs*, 159, 1-107 .

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998).The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634–644.

Roeder, S., & Paulhus, D. L. (2009, February). Measuring consumer knowledge in the face of exaggeration and sabotage. *Poster presented at the meeting of the Society for Consumer Psychology*, San Diego.

Schmit, M., & Ryan, A. (1993). The Big-Five in personnel selection: Factor structure in applicant and non-applicant populations. *Journal of applied psychology*, *78*, 966-974.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*, 607-620.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, *87*, 211-219.

Stokes, G. S., Hogan, J. B., & Snell, A. F. (1993). Comparability of incumbent and applicant samples for the development of biodata keys: The influence of social desirability. *Personnel Psychology*, *46*, 739-762.

Swets, J. A. (1964). *Signal detection and recognition by human observers.* New York: Wiley.

Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity*, 8, 196-213.

Tonković, M., Galić, Z., &Jerneić, Ž. (2011).The construct validity of over-claiming as a measure of egoistic enhancement. *Review of Psychology, 18,* 13-21.

Uziel, L. (2010). Rethinking social desirability scales from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243-262.

Uziel, L. (2014). Impression management ("lie") scales are associated with interpersonally oriented self- control, not other- deception. *Journal of personality, 82*, 200-212.

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

White, L. A., Young, M. C., & Rumsey, M. G. (2001). ABLE implementation issues and related research. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 525-558). Hillsdale, MJ: Erlbaum.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23,* 419-427.

Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. Educational and Psychological Measurement, 24, 551-562.

Williams, K. M., Paulhus, D. L., & Nathanson, C. (2002, August). The nature of over-claiming: Personality and cognitive factors. In *Poster presented at the 110th annual meeting of the American psychological association, Chicago, IL [August, 2002].*

Zerbe, W. J., &Paulhus, D.L. (1987). Social desirable responding in organizational behavior: *A Reconception. Academy of Management Review*, *12*, 250-264.

Ziegler, M., Kemper, C., & Rammstedt, B. (2013). The Vocabulary and Overclaiming Test (VOC-T). *Journal of Individual Differences*, *34*, 32-40.

Ziegler, M., MacCann, C., Roberts, R. D. (2012). *New Perspectives on Faking in Personality Assessment*. Oxford University Press.

MEASURES

BASELINE ASSESSMENT (Time 1)

Instructions: Please read the following statements and indicate the degree to which each statement is an accurate description of you with respect to how you behave. Describe yourself as you generally are, not as you wish to behave in the future. Describe yourself as you honestly see yourself in relation to other others who are of the same sex as you are, and roughly your same age. Your responses will be kept in absolute confidence, so please respond honestly. Please response using the following scale:

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

AGREEABLENESS
1. I insult people.**
2. I am interested in people.
3. I feel little concern for others.**
4. I have a soft heart.
5. I am hard to get to know.**
6. I feel others' emotions.
7. I inquire about others' well-being.
8. I love children.
9. I have a good word for everyone.
10. I think of others first.

ARCHIEVEMENT STRIVING
11. When it comes to class participation and course work, I am easy-going and lackadaisical.**
12. When I start a self-improvement program to improve my study habits, I usually let it slide after a few days.**
13. I strive for excellence in everything I do at school.
14. I'm something of a "workaholic" at school.

CONSCIENTIOUSNESS
15. I am always prepared.
16. I leave my belongings around.**
17. I get chores done right away.
18. I often forget to put things back in their proper place.**
19. I follow a schedule.

20. I neglect my duties.**
21. I do things according to a plan.
22. I do things in a half-way manner.**
23. I make plans and stick to them.
24. I leave a mess in my room.**

COMPETITIVENESS
25. I'm known for my prudence and common sense when it comes to academic activities.
26. I often come into academic situations (e.g., tests, group projects, etc.) without being fully prepared.**
27. I don't seem to be completely successful at anything school-related.**
28. I am efficient and effective at my schoolwork.

EXTRAVERSION
29. I am the life of the party.
30. I often feel uncomfortable around others.**
31. I start conversations.
32. I keep in the background.**
33. I don't like to draw attention to myself.**
34. I don't mind being the center of attention.
35. I take charge.
36. I find it difficult to approach others.**
37. I feel at ease with people.
38. I bottle up my feelings.**

EMOTIONAL STABILITY
39. I get stressed out easily.**
40. I am relaxed most of the time.
41. I am easily disturbed.**
42. I am not easily bothered by things.
43. I change my mood a lot.**
44. I seldom get mad.
45. I get irritated easily.**
46. I get angry easily.**
47. I feel threatened easily.**
48. I take offense easily.**

OPENNESS
49. I have a rich vocabulary.
50. I have difficulty understanding abstract ideas. **
51. I have excellent ideas.
52. I do not have a good imagination.**

53. I use difficult words.
54. I have difficulty imagining things.**
55. I am full of ideas.
56. I will not probe deeply into a subject.**
57. I catch on to things quickly.
58. I am good at many things.

SELF-DECIPLINARY
59. I'm pretty good about pacing myself so as to get assignments done on time.
60. When I study, I waste a lot of time before settling down to work.**
61. Once I start a class project, I almost always finish it.
62. When a term project gets too difficult, I'm inclined to start a new one.**

Note: Items indicating ** are reverse coded.

DIVERSION TASK (Time 2, for both Motivated Condition and Control Condition)

Resume Content Analysis

Introduction: The following questions are designed to elicit information about how you evaluate resume content. Each question will present you with information commonly found on a resume. You will be asked to judge how that content represents specific attributes including individual abilities, work style characteristics, and skills. Each page asks you to do the same thing, but for different resume content. The resume content of interest will appear in bold face type in each question.

1	2	3	4
Not at all	Somewhat	Moderate	Considerable

1. When listed on a resume, to what degree does having earned **an associate's degree** mean that the applicant has the following **abilities**?

____Verbal Ability-the ability to comprehend spoken and written words, and communicate information so that others will understand

____Idea Generation and Reasoning Ability--the ability to generate ideas, identify and solve problems, and use reason and logic to organize information.

____Quantitative Ability--the ability to do mathematical operations, and understand and solve math related problems.

____Memory--the ability to remember information such as words, numbers, pictures and procedures

____Perceptual Ability--the ability to quickly make sense of and organize information, and identify and compare patterns, objects, or numbers

____Spatial Ability -- the ability to know one's location in relation to the environment, and imagine how something will look after it is rearranged or moved

____Attentiveness -- the ability to concentrate and not be distracted while preforming a task,

and to efficiently shift back and forth between activities

____Practical Intelligence -- the ability to generate useful ideas, think things through logically,innovate, and think analytically

2. When listed on a resume, to what degree does having earned **an associate's degree** mean that the applicant has the following **work styles**?

____Achievement orientation-- setting and seeking to attain personal goals, and striving to be competent in one's work.

____Social influence -- having the capacity to impact others, and display energy and leadership

____Interpersonal Orientation -- being pleasant, cooperative, and sensitive to others

____Adjustment -- being mature, poised, and flexible, and having the capacity to cope with pressure, stress, and criticism

____Conscientiousness -- being dependable, trustworthy, accountable, and attentive to details

3. When listed on a resume, to what degree does having earned **an associate's degree** mean that the applicant has the following **skills**?

____Technical skills -- skills in areas such as design, operations analysis, programming, operating and maintaining equipment, and troubleshooting.

____Job knowledges -- awareness of a specific work context or occupation

____Resources Management Skills -- skill in managing time, financial resources, material resources, and personnel resources

____Social skills -- skill in coordinating, persuading, negotiating, and instructing others, having social perceptiveness and a service orientation

____System skills -- skill in areas of system analysis including envisioning and positioning, initiating chance, and considering long-term outcomes

4. When listed on a resume, to what degree does having earned **a Bachelor's degree** mean that the applicant has the following **abilities**/**work styles**/ **skills**?

5. When listed on a resume, to what degree does having earned **a Master's degree** mean that the applicant has the following **abilities**/**work styles**/ **skills**?

6. When listed on a resume, to what degree does having earned **a degree in a job-related field of study** mean that the applicant has the following **abilities**/**work styles**/ **skills**?

7. When listed on a resume, to what degree does having earned **a grade point average (GPA) of 3.0 or better in a job-related field of study** mean that the applicant has the following **abilities**/**work styles**/ **skills**?

8.When listed on a resume, to what degree does having earned **a professional certification [e.g.**

Professional in Human Resources (PHR)] mean that the applicant has the following **abilities/work styles/ skills**?

9. When listed on a resume, to what degree does **participation in an international study program (e.g. study abroad)** mean that the applicant has the following **abilities/work styles/ skills**?

10. When listed on a resume, to what degree does **fluency in a foreign language** mean that the applicant has the following **abilities/work styles/ skills**?

11. When listed on a resume, to what degree does **expertise with software applications** mean that the applicant has the following **abilities/work styles/ skills**?

12. When listed on a resume, to what degree does having an **internship experience** mean that the applicant has the following **abilities/work styles/ skills**?

13. When listed on a resume, to what degree does having **non-job related work experience** mean that the applicant has the following **abilities/work styles/ skills**?

14. When listed on a resume, to what degree does having **non-job related work experience** mean that the applicant has the following **abilities**?

Note: Items of Question 4 to Question 14 are exactly the same as items of questions 1, 2, and 3.

PRE-EMPLOYMENT ASSESSMENT (Time 2, Motivated Condition)

PERSONALITY INVENTORY

Instructions: In this inventory you will find a series of statements which a person might use to describe himself or herself. Read each statement carefully and indicate the extent to which each statement below describes the ways you typically think, feel, and act in your daily life. Please response using the following scale:

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

AGREEABLNESS
  2. I am not interested in other people's problems.**
 17. I am not really interested in others.**
 31. I am indifferent to the feelings of others.**
  9. I sympathize with others' feelings.
 24. I take time out for others.

117

38. I make people feel at ease.

45. I know how to comfort others.

52. I am on good terms with nearly everyone.

59. I show my gratitude.

66. I love to help others.

ACHIEVEMENT STRIVING

75. For school projects, I have a clear set of goals and work toward them in an orderly fashion.

82. I work hard to accomplish my academic goals (e.g., completing homework, making grades, etc).

72. In my college studies, I don't feel like I'm driven to get ahead.**

77. I strive to achieve all I can at college.

COUNSCIENTIOUSNESS

11. I make a mess of things.**

25. I shirk my duties.**

39. I waste my time.**

54. I find it difficult to get down to work.**

4. I pay attention to details.

18. I like order.

32. I am exacting in my work.

47. I continue until everything is perfect.

61. I love order and regularity.

67. I like to tidy up.

COMPETITIVENESS

76. I don't take student duties like completing professor evaluations very seriously.**

79. I keep myself informed and usually make intelligent decisions on class-related projects (e.g., long-range assignments, term papers, group projects, etc.).

81. I pride myself on my sound judgment in my college class work.

73. I'm a very competent person when it comes to college level work.

EXTRAVERSION

8. I have little to say.**

22. I am quiet around strangers.**

36. I don't talk a lot.**

51. I am a very private person.**

58. I wait for others to lead the way.**

1. I feel comfortable around people.

15. I talk to a lot of different people at parties.
30. I make friends easily.
44. I know how to captivate people.
65. I am skilled in handling social situations.

EMOTIONAL STABILITY
5. I worry about things.**
19. I get upset easily.**
34. I have frequent mood swings.**
41. I often feel blue.**
48. I panic easily.**
55. I get overwhelmed by emotions.**
62. I get caught up in my problems.**
69. I grumble about things.**
12. I seldom feel blue.
27. I rarely get irritated.

OPENNESS
14. I am not interested in abstract ideas. **
28. I try to avoid complex people.**
42. I avoid difficult reading material.**
7. I have a vivid imagination.
21. I am quick to understand things.
35. I spend time reflecting on things.
49. I carry the conversation to a higher level.
57. I can handle a lot of information.
64. I love to read challenging material.
70. I love to think up new ways of doing things.

IMPRESSION MANAGEMENT
3. I am sometimes irritated by people who ask favors of me.**
6. I sometimes try to get even, rather than forgive and forget.**
16. Sometimes at elections I vote for candidates I know little about.**
46. Once in a while I laugh at a dirty joke.**
63. There have been occasions when I have taken advantage of someone.**
13. I am always courteous, even to people who are disagreeable.
40. When I take sick-leave from work or school, I am always as sick as I say I am.
43. I always apologize to others for my mistakes.
60. I always tell the truth.
68. I have never dropped litter on the street.

SELF-DECEPTIVE ENHANCEMENT

20. I worry quite a bit over possible misfortunes.**
23. Life is a strain for me most of the time.**
26. I have not always been honest with myself.**
29. In a group of people I have trouble thinking of the right things to talk about.**
50. I have several times given up doing something because I thought too little of my ability.**
56. I have sometimes thought that my parents hated me.**
10. I never regret my decisions.
33. I rarely appreciate criticism.**
37. People often disappoint me.**
53. I have thought of committing suicide in order to get back at someone.

SELF-DICIPLINARY

78. In college, I am a productive person who always gets the work done.
80. I have trouble making myself do what I should for classes.**
74. There are so many little class assignments that need to be done that I sometimes just ignore them all.**
71. I have a lot of self-discipline in my course work.

Note: Items indicating ** are reverse coded.

TASK EXPERIENCE SURVEY

Instructions: Listed below are a series of tasks related to clerical work, computer programming and use, research skills, and interpersonal relationships, most of which you will perform at the position at InSat Corporation's Summer Internship Program. Please indicate how frequently you have done each task during the last 12 months, using the following scale:

1 Never
2 A couple of times
3 Quite a number of times
4 Often

1. Utilize the Murray-Web system to locate unpublished research articles. (BS)
2. Operate a fax machine.
3. Arrange/set up tables for banquets and events.
4. Operate a coping machine.
5. Organize files using the FolderPro Filing Scheme. (BS)
6. Type at least 70 wpm on a QWERTY keyboard.
7. Plan and host a group meeting.

8. Use Johnson's Dyadic Approach of avoiding conflict in work teams. (BS)
9. Manage databases using Microsoft Access.
10. Assess the feasibility of a business project using the Expert Assessment. (BS)
11. Format a professional paper using the APA format.
12. Create a presentation using Microsoft PowerPoint.
13. Use HURIER skills to present ideas to coworkers or clients.
14. Edit digital photographs using Adobe Photoshop.
15. Use Rastmen's Power theory to motivate others in your team. (BS)
16. Video chat with friends using an iPad.
17. Create slideshow using Picslide program. (BS)
18. Book flights on the Internet.
19. Set up a multi-party online chat with Skype.
20. Create charts and tables using Microsoft Excel.
21. Use the silvernman leadership skill to distribute teamwork appropriately. (BS)
22. Use the Outlook online calendar.
23. Utilize the Web of Science to locate published research articles.
24. Operate a Nephogram projector. (BS)
25. Use the "foot-in-the-door" technique to persuade potential buyers.

Note: Items indicating (BS) represent bogus items.

GENERAL KNOWLEDGE SURVEY

Instructions: In the following list of 25 items you will see various historical figures, names, books, authors, and various terms. Please rate your familiarity with each item by clicking the button beneath your selected response on the following scale:

Never heard of it	Slightly familiar	Somewhat familiar	Familiar	Very familiar
o	o	o	o	o

For example, if the item said "Bill Clinton", you would probably click the button beneath "Familiar" (or "Very familiar") to indicate that you are familiar with who he is.

Never heard of it	Slightly familiar	Somewhat familiar	Familiar	Very familiar
o	o	o	●	o

However, if the item said "Fred Gruneberg", you would probably click the button beneath "Never heard of it" because you have never heard of him, or you might click the button beneath "Slightly familiar" because he sounds vaguely familiar.

Never heard of it	Slightly familiar	Somewhat familiar	Familiar	Very familiar
●	o	o	o	o

Once you have read and understood these instructions then please feel free to begin responding to the list of items below.

1   Houdini

2   Charlotte Bronte

3   meta-toxins (F)

4   myth

5   Antigone

6   chlorarine (F)

7   alliteration

8   Gail Brennan (F)

9   Queen Shattuck (F)

10   Lewis Carroll

11   free will

12   Dale Carnegie

13   Murphy's Last Ride (F)

14   sentence stigma (F)

15   Bay of Pigs

16   hyperbole

17   The Aeneid

18   euhpamism

19   double entendre

20   consumer apparatus (F)

21   blank verse

22   shunt-word (F)

23   art deco

24   Artemis

25   a capella

Note: Items indicating (F) represent foils

RESEARCH INVENTORY (Time 2, Control Condition)

SURVEY 1

Instruction: In this inventory you will find a series of statements which a person might use to describe himself or herself. Read each statement carefully and indicate the extent to which each statement below describes the ways you typically think, feel, and act in your daily life. Please response using the following scale:

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

(Items of Survey1 are exactly the same as Personality Inventory, Time 2 at Motivated-Condition)

SURVEY 2

Instructions: Listed below are a series of tasks related to clerical work, computer programming and use, research skills, and interpersonal relationships. You may or may not have previous experience with them. Please indicate how frequently you have done each task during the last 12 months, using the following scale:

1 Never
2 A couple of times
3 Quite a number of times
4 Often

(Items of Survey2 are exactly the same as items of Task Experience Survey, Time 2 at Motivated-Condition)

SURVEY 3

Instructions: In the following list of 25 items you will see various historical figures, names, books, authors, and various terms. Please rate your familiarity with each item by clicking the button beneath your selected response on the following scale:

Never heard of it   Slightly familiar Somewhat familiar   Familiar   Very familiar
         o   o   o   o   o

For example, if the item said "Bill Clinton", you would probably click the button beneath "Familiar" (or "Very familiar") to indicate that you are familiar with who he is.

        Never heard of it   Slightly familiar Somewhat familiar Familiar   Very familiar
               o   o   o   ●   o

However, if the item said "Fred Gruneberg", you would probably click the button beneath "Never heard of it" because you have never heard of him, or you might click the button beneath "Slightly familiar" because he sounds vaguely familiar.

        Never heard of it   Slightly familiar Somewhat familiar Familiar   Very familiar
               ●   o   o   o   o

Once you have read and understood these instructions then please feel free to begin responding to the list of items below.

(Items of Survey3 are exactly the same as the items of General Knowledge Survey, Time 2 at Motivated-Condition)

SURVEY INSTRUCTIONS AT TIME 2

FOR TIME 2, MOTIVATED CONDITION

Thank you for your willingness to help us develop the pre-screening assessment for our paid summer internship program.

InSat Corporation's Summer Internship Program offers college students excellent development opportunities involving real-world job experience. In the program, your duties include both basic clerical tasks and business related tasks. Some examples of the former tasks include working with Microsoft office software, sending emails, organizing files and taking memos for meetings. Examples of the latter tasks include selling ideas to clients, working in a team to come up business plans, and conduct basic level data analysis.

When screening for the Summer Internship Program, we are interested in selecting college students from a variety of backgrounds who are hardworking and detail-oriented, sociable and able to work well with others, bright and open-minded, and able to tolerate stress.

At the end of the questions, you'll be asked to indicate if you would like us to contact you about a summer intern position. If you select "yes," we'll take a close look at your responses to these questions and evaluate them given our needs.

FOR TIME 2, CONTROL CONDITION

Now you will be completing three surveys we mentioned earlier. When completing these surveys, please be as honest as you can be. Thus, do your best to respond to every statement in a genuine and honest way.

Appendix C

DECEPTION SCRIPT (TIME 2, MOTIVATED CONDITION)

Dr. Fan:

Hello, I am Dr. Fan. Thanks for coming to my lab and completing the first part of the study, which is my research. For the second part, you are going to do something completely different, and that is not even my research, as I am helping a friend of mine, who is a CEO for a company called InSat Corporation. They are developing a pre-screening assessment for recruiting and selecting college students into their paid summer internship program. InSat needs a collection of college students to answer the questions to help with development. We have agreed to help them gather the data needed. This pre-screening assessment consists of a personality inventory, a basic skills survey, and a general knowledge survey. It takes about 20 minutes to complete. Based on this pilot test, InSat will revise and finalize the pre-screening assessment. They plan to start a large-scale campus recruitment campaign in the spring semester of 2014 at several major universities in the Southeast U.S. This of course includes Auburn University.

When the CEO of InSat approached me for help, I told him that we have this wonderful SONA system through which we are able to get enough college students to pilot-test their pre-screening assessment. However, I also asked him if they could offer something additional? They came up with two additional benefits. First, InSat has agreed to sponsor the drawing for ten $50 cash rewards. We will do the drawing at the end of the semester, and there are 10 lucky participants who will each receive $50 cash reward from InSat.

As the CEO is talking to me, he is like, "You know what? Our campus recruitment campaign will cover Auburn University. Maybe some of the participants are interested in our paid summer internship program. If they indicate that they are interested in the summer internship program by checking the "Yes" box at the end of the assessment, we will review their answers more carefully, and if they perform well on the assessment, we are happy to give them the earlier consideration for the summer internship program." However, no need to feel pressured. That is, taking the test does not necessarily mean that you are required to attend the summer internship program. InSat has written up an introduction to their paid summer internship program and you can find more details about it on the first page of the survey.

So, now I'm going to have you complete the pre-screening assessment that InSat is developing (password: InsatTest2). Please take your Stage and respond carefully. Now you may open the survey link on the computer, login, and start the assessment.

DEBRIEFING SCRIPTS

DEBRIEFING SCRIPT FOR MOTIVATED CONDITION

Thanks very much for your participation in this study. But I have to disappoint you—we are actually not working with InSat Corporation. InSat Corporation is actually not existent. I apologize for the deception and the disappointment you might have.

But let me explain—the deception condition was created for research purposes. The aim of this study is to examine faking under an application condition. As you know, applicants tend to enhance themselves and respond in a more socially desirable way when applying for a job, a phenomenon called "social desirability" or "faking." There is some evidence that faking may render the selection questionnaires invalid. We want to study faking, and one thing we have to do is to simulate an application situation, under which applicants are likely to engage in faking when completing the pre-employment assessment. In order to simulate an application situation, we pretended to be offering a potential job opportunity that you might have been interested in. We will compare your responses to a control group--another group of students who did the same tasks as you guys but were not offered the potential job opportunity. As you may notice, the first task you did (the resume analysis task), is a diversion task which is not related to the research purpose. The research is supported by Dr. Fan's (my advisor) research fund and the drawing for the 50$ cash reward is real. We have ten$50 cash rewards to give out to participants including students in this condition and in the control condition. We will do the drawing by the end of this semester. Each participant will get an equal chance to win the cash rewards.

Also, as mentioned in the informed consent, in order to validate the questionnaire, we need to obtain your GPA, ACT and/or SAT scores from the University Registrar. In addition, we also need two of your friends to evaluate your adjustment at college. We would prefer someone who knows a lot about your college life, such as close friends, roommates, classmates, etc. If they participate, they will also get a chance to win one of twenty $15 cash rewards. These rewards are separate from yours so their participation will not influence your chance to win the cash reward. In addition, each completed peer-survey will earn you .5 hour of SONA credit. The

contact information of your peers and their responses to the peer-survey will be kept confidential, and the information they provide will only be used for research purposes.

Given that this study involves a deception, you have the choice of not allowing us to use your data in further analysis. If you choose this option, please notify the experimenter now; you will also be excused from the peer-rating survey part of the study. In this case, you will still receive 1.5 hour of SONA credit and be entered into the cash drawing. Otherwise, we need you to send an email to your peers right now stating that you need their help for this research, and the researchers will follow up with them via email.

An important note: Since the data collection is currently under way and we have more participants coming in to do the study during the rest of the semester, we ask that you not share the study details (particularly the deception part) with other students even if they will not participate in this study. This way the integrity of the study will be protected. Thank you.

DEBRIEFING SCRIPT FOR CONTROL CONDITION

Thanks very much for your participation. What you have just done is actually one part of a research study examining applicant faking. As you know, applicants tend to enhance themselves and respond in a more socially desirable way when applying for a job, a phenomenon called "social desirability" or "faking." There is some evidence that faking may render the selection questionnaires invalid.

We have two groups in this research. In the experimental group, participants were led to believe that they were being considered for a potential job opportunity while completing the surveys. This should prompt these participants to respond in a more socially desirable style. In the control group, participants (you guys) are told to complete the questionnaire for research purpose, and thus you presumably have no desire to fake. As you may notice, the first task you did (the resume analysis task) is a diversion task which is not related to the research purpose. This research is supported by Dr. Fan's (my advisor) research fund and the drawing for the 50$ cash reward is real. We have ten 50$ cash rewards to give out to participants including students in this condition and in the experiment condition. We will do the drawing by the end of this semester. Each participant will get an equal chance to win cash rewards.

Also, as mentioned in the informed consent, in order to validate the questionnaire, we need to obtain your GPA, ACT and/or SAT from the University Registrar. In addition, we also need two of your friends to provide their information about you. We would prefer someone who knows a lot about your college life, such as close friends, roommates, classmates, etc. If they participate, they will also get a chance to win one of twenty $15 cash rewards. These rewards are separate from yours so their participation will not influence your chance to win the cash reward. In addition, each completed peer-survey will earn you .5 hour of SONA credit. The contact information of your peers and their responses to the peer-survey will be kept confidential, and the information they provide will only be used for research purposes.

Now we need you to send an email to your peers right now stating that you need their help for this research, and the researchers will follow up with them via email.

An important note: Since the data collection is currently under way and we have more participants coming to do the study during the rest of the semester, we ask that you not share the study details (particularly the deception part) with other students even if they will not participate in this study. This way the integrity of the study will be protected. Thanks.