Applicant Faking, Social Desirability, and the Prediction of Counterproductive Work Behaviors

Mitchell H. Peterson, Richard L. Griffith, and Joshua A. Isaacson

*Florida Institute of Technology*

Matthew S. O'Connell

*Select International Inc.*

Phillip M. Mangos

*Kronos Inc.*

Recent studies have pointed to within-subjects designs as an especially effective tool for gauging the occurrence of faking behavior in applicant samples. The current study utilized a within-subjects design and data from a sample of job applicants to compare estimates of faking via within-subjects score change to estimates based on a social desirability scale. In addition, we examined the impact of faking on the relationship between Conscientiousness and counterproductive work behaviors (CWBs), as well as the direct linkage between faking and CWBs. Our results suggest that social desirability scales are poor indicators of within-subjects score change, and applicant faking is both related to CWBs and has a negative impact on the criterion-related validity of Conscientiousness as a predictor of CWBs.

For more than 15 years, meta-analytic efforts aimed at examining personality–performance relationships (e.g., Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991) have been cited as evidence of the utility of personality inventories. Although recent work has questioned the validity of personality measures (e.g., Morgeson et al., 2007), particularly by noting significant concern over the use of self-report assessments, this criticism has spurred strong rebuttals outlining the value of personality variables in organizations (e.g., Ones, Dilchert, Viswesvaran, & Judge, 2007; Tett & Christiansen, 2007). One of the controversial issues related to personality testing has been the pervasive concern over the potential for job applicants to provide exaggerated or distorted responses to self-report personality assessments (Griffith & McDaniel, 2006; Schmitt & Oswald, 2006). This phenomenon has typically been referred to as *applicant faking*. Although the body of faking literature has grown considerably in recent years, methodological challenges may have contributed to the slow progress of research examining the issue.
From the earliest studies on faking to the most current approaches, researchers have strived to develop a methodology that is capable of capturing and explaining this complex behavior. This struggle is apparent given the large number of strategies employed to study faking that can be found in the extant literature. These methodologies have included examinations of directed or motivated faking in the laboratory (e.g., McFarland & Ryan, 2000), between-subjects investigations of applicant-incumbent differences (e.g., Hough, Eaton, Dunnette, Kamp, & McClory, 1990), social desirability scales (e.g., Ones, Viswesvaran, & Reiss, 1996), self-reported faking (e.g., Donovan, Dwight, & Hurtz, 2003), the use of deception to simulate an applicant setting (e.g., Griffith, Malm, English, Yoshita, & Gujar, 2006), Monte Carlo simulations (e.g., Converse, Peterson, & Griffith, 2009; Komar, Brown, Komar, & Robie, 2008; Schmitt & Oswald, 2006), and finally, within-subjects designs with applicant samples (e.g., Arthur, Glaze, Villado, & Taylor, 2010; Ellingson, Sackett, & Connelly, 2007; Griffith, Chmielowski, & Yoshita, 2007; Hogan, Barrett, & Hogan, 2007). The study of faking behavior has also been extended to biodata questionnaires, with researchers comparing the responses of job applicants to samples of participants instructed to fake-good or respond candidly (e.g., Becker & Colquitt, 1992). The widely varying methods employed in the study of faking behavior have nonetheless left the body of literature without a conclusive answer to questions regarding the prevalence and personnel selection-related impact of faking behavior.

Recent work has examined faking behavior of actual job applicants via examinations of within-subjects score change across conditions that should vary in their motivational influences on respondent behavior (Arthur et al., 2010, Ellingson et al., 2007; Griffith et al., 2007; Hogan et al., 2007). Despite these studies’ strengths, they have by no means offered an unequivocal answer to the questions surrounding the faking phenomenon. Our goal in the present study, therefore, was to attempt to extend the findings of these studies by using a within-subjects investigation of faking in real job applicants. In addition, we attempt to highlight the key differences in conclusions regarding the occurrence and impact of faking behavior when different methodologies are employed to study the phenomenon.

**WITHIN-SUBJECTS DESIGNS USING APPLICANT DATA**

The use of within-subjects methodologies in the study of faking behavior offers several benefits. Most important, within-subjects designs allow the researcher to directly observe individual-level score changes on personality scales across assessment contexts (e.g., applicant vs. research) rather than measuring additional variables (e.g., social desirability, bogus item endorsement) and using those measurements to make inferences regarding the occurrence of faking. In addition, within-subjects designs allow for the identification of individuals engaging in faking behavior using confidence interval methodologies that account for the measurement error inherent in personality assessments (e.g., Griffith et al., 2007). A real-world within-subjects applicant faking design allows researchers to collect data in a setting where respondents may be naturally inclined to respond in a socially desirable manner (Ellingson, Sackett, & Hough, 1999). In addition, from a measurement perspective, this design provides a mechanism to analyze the consistency of responses as well as maximizing the power of each observation (D. T. Campbell & Stanley, 1963). Ryan and Boyce (2006) noted that this type of a design represents the “gold-standard” (p. 363) of faking research designs.
Several within-subjects studies have recently appeared in the faking literature; however, they have not resulted in an unequivocal answer to the question of whether faking is a common occurrence in real selection settings (Arthur et al., 2010, Ellingson et al., 2007; Griffith et al., 2007; Hogan et al., 2007). First, Ellingson et al. (2007) used a within-subjects design to examine score change across selection and personal development contexts. Using a large archival data set, the authors identified individuals who had completed a personality assessment on two occasions across four specific types of test–retest conditions (i.e., development–development, development–selection, selection–development, selection–selection). Ellingson et al. (2007) argued that score change across the two contexts was subject to a variety of influences (beyond faking), including personality change due to developmental feedback and personality change over time. In an attempt to isolate the effects of faking on score change (which, as the authors noted, should only occur in the development–selection or selection–development conditions), the authors controlled for potential true personality change over time and potential change due to developmental feedback. In the end, this procedure resulted in a negligible effect size for faking across contexts (\(d = .075\)) when effect sizes were averaged across the 18 California Psychological Inventory (CPI) subscales. However, effect sizes for single personality scales within the full assessment were as high as .64 for the development–selection context, with several scales demonstrating effect sizes greater than .40.

In another within-subjects investigation, Hogan et al. (2007) used archival data to examine personality change in a sample of individuals who retook an assessment after being denied employment. Hogan et al. suggested that this represented a condition in which individuals should have been motivated to engage in distortion in order to improve their scores on the assessment. Like Ellingson et al. (2007), Hogan et al. also found little evidence of faking. The authors noted that only a small portion of the sample (anywhere from 1.7% to 5.2% of applicants across scales) significantly raised their scores on the second assessment, with only .06% doing so on all five personality scales. In addition, due to evidence suggesting that score change was normally distributed, with a mean of zero, the authors argued that most score changes were the result of random measurement error.

In contrast to the findings of Ellingson et al. (2007) and Hogan et al. (2007), recent work by Griffith et al. (2007) and Arthur et al. (2010) found considerable evidence of faking in applicant samples. Griffith et al. (2007) used a within-subjects design to examine faking (via score change from applicant to honest assessments) in a sample of applicants to a temporary employment agency. The authors reported significant mean-level score differences between applicant and honest responses, in addition to finding that between 22% and 49% of applicants faked their responses (depending upon how the faking variable was operationalized). In addition, the authors noted changes in simulated individual hiring decisions across the applicant and research contexts. Although the comparison of responses across applicant and research contexts represented an effective measurement of faking, the small and unique (temporary employees) sample may limit the generalizability of the Griffith et al. (2007) study.

Arthur et al. (2010) examined the prevalence of applicant faking in an unproctored internet testing selection setting across two studies. The authors gathered personality test responses from a sample of job applicants who were subsequently contacted (typically after more than 1 year) and asked to take part in a research study using the same measure of personality. Across both studies, Arthur et al. reported significant mean-level differences between applicant and research context scores (with applicant condition scores being higher) and percentages of individuals identified as
having significantly elevated applicant scores (i.e., applicant scores exceeding the upper bound of a confidence interval formed around the research context score) ranging from 15 to 36% across the Big Five dimensions.

Although the prevalence of within-subjects investigations of faking behavior in samples of real job applicants has recently increased, the discord in findings across these studies suggests that more thorough investigations using similar methodologies are warranted. Furthermore, given the methodological challenges inherent in carrying out a within-subjects investigation of faking behavior with job applicants, additional research employing such designs has the potential to add not only to our understanding of the nature of faking behavior but also to spur new methodological innovations on the part of researchers. In the present investigation, we sought to extend the findings of the aforementioned studies by examining the prevalence and correlates of applicant faking using a within-subjects design with real job applicants. Most directly, this study extends the methodologies employed by Arthur et al. (2010) and Griffith et al. (2007) by gathering data from participants across job applicant and research contexts.

We believe that the present study offers a useful extension and contribution to the literature in that we also examined correlates of faking behavior from three perspectives. First, we set out to investigate the degree to which a measure of social desirability (SD) was sensitive to actual score elevation in the applicant context. Next, we attempted to determine whether the extent to which individuals engage in faking behavior was associated with other negative workplace behaviors, namely, self-reported counterproductive work behaviors (CWBs). Finally, we examined the degree to which faking resulted in decrements to the validity of Conscientiousness as a predictor of self-reported CWB.

**THE CURRENT STUDY**

The present study represents an extension of the works by Arthur et al. (2010) and Griffith et al. (2007) in that job applicants completed an assessment of Conscientiousness during an employment screening process and completed a second administration of the assessment in a research context. Based on fact that the present study used similar methodology to Arthur et al. and Griffith et al. (2007), we expected to replicate the findings of these earlier works. Therefore, we hypothesized the following:

**H1:** Conscientiousness scores obtained in the applicant context will be significantly higher than Conscientiousness scores from a research administration of the same assessment.

**H2:** A significant number of individuals will be identified as having faked the Conscientiousness measure in the applicant condition, such that their applicant scores exceed the upper bound of a 95% confidence interval surrounding their score from the research context.

**SD Scores as Indicators of Applicant Faking**

Many faking studies have used SD scales as indicators of faking behavior (e.g., Hough et al., 1990; Ones et al., 1996). These studies have frequently been cited as evidence that faking does not represent a legitimate concern for organizations using personality assessments (e.g., Ones
et al., 1996). However, the use of SD scales to identify or correct for response distortion has been questioned in terms of both its methodological soundness (e.g., Burns & Christiansen, 2006; Griffith & Peterson, 2008; Smith & Ellingson, 2002) and its usefulness in improving selection outcomes (e.g., Schmitt & Oswald, 2006). Specifically, researchers have expressed concern over whether SD scales are assessing individual variation in response styles, or substantive personality variance. Smith and Ellingson (2002) found that SD demonstrated consistent relationships with the Big Five traits of Conscientiousness and neuroticism. In addition, several authors have noted that individuals may attempt to tailor their responses to multidimensional personality inventories to fit the requirements of specific positions (e.g., Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Mahar et al., 2006). For example, Becker and Colquitt (1992) found that the most job relevant items on a biodata questionnaire tended to be most prone to faking by applicants. If this is in fact the case, then the use of SD scales in practice is questionable due to the fact that this method assumes that individuals are faking to a similar degree across all scales of the personality inventory.

Although evidence gathered from directed faking manipulations (e.g., Hough et al., 1990) suggests that SD scales are associated with response distortion, this finding has yet to be consistently replicated in within-subjects investigations of faking using actual applicant samples. An additional within-subjects investigation of faking carried out by Griffith et al. (2006) provided little support for the validity of SD scales as indicators of faking. Griffith et al. reported correlations between applicant faking on the Conscientiousness scale of the NEO-Five Factor Inventory (Costa & McCrae, 1992) and scores on the Impression Management and Self-Deceptive Enhancement subscales of the Balanced Inventory of Desirable Responding (Paulhus, 1998) of .11 and .12 (both nonsignificant), respectively. Griffith et al. (2006), however, used a simulated applicant setting in which students were led to believe they were applying for a job, rather than data gathered from real job applicants.

There appears to be little evidence of a relationship between applicant faking and SD scores. Therefore, we hypothesized the following with regard to SD measures as indicators of faking behavior:

H3: Scores on a measure of SD will not be significantly related to applicant faking on the Conscientiousness measure when operationalized as score change from the applicant to research contexts.

H4: Classifications of individual respondents suspected of faking using a measure of SD will be statistically independent of classifications using significant within-subjects score change.

Applicant Faking and the Prediction of CWBs

Previous investigations examining faking as a potential threat to the validity of personality variables as predictors of job performance have offered widely varied conclusions. Studies using measures of social desirability (e.g., Barrick & Mount, 1996; Hough et al., 1990; Ones et al., 1996) as indicators of faking behavior, or simulating the impact of corrections for social desirability (e.g., Schmitt & Oswald, 2006) have typically found that faking has little influence on the criterion-related validity of personality assessments. In contrast, additional studies using directed faking manipulations (e.g., Mueller-Hanson, Heggestad, & Thornton, 2003) and Monte Carlo
simulation (e.g., Converse et al., 2009; Komar et al., 2008) have reported findings suggesting that faking has the potential to negatively impact selection outcomes. The studies just mentioned, however, devoted little attention to the variety of performance criteria that personality variables may predict.

As a meta-analysis by Hogan and Holland (2003) demonstrated, the greatest predictive potential for personality variables may likely be in the prediction of specific or narrow performance criteria. With several multidimensional conceptualizations of the job performance domain gaining widespread acceptance from researchers (e.g., J. P. Campbell, McCloy, Oppler, & Sager, 1993; Motowidlo, Borman, & Schmit, 1997), personality researchers have an array of potential criterion variables to work with. Furthermore, this call for a broader conceptualization of the performance domain has also been extended to examinations of applicant faking (Peterson & Griffith, 2006).

CWBs represent one specific aspect of organizational behavior that has demonstrated consistent relationships with personality variables. These behaviors typically encompass actions that go against organizational norms and have the potential to negatively impact both the organization and the individuals comprising it (Berry, Ones, & Sackett, 2007). Two recent meta-analyses (Berry et al., 2007; Salgado, 2002) have identified the personality variable of interest in the present study, Conscientiousness, as a consistent predictor of CWBs (i.e., conscientious individuals are less likely to engage in CWBs).

Faking has the potential to both play a role in the ability of personality variables, like conscientiousness, to predict CWBs and functioning as a predictor of counterproductive behavior in its own right. First, as the meta-analyses by Berry et al. (2007) and Salgado (2002) described, Conscientiousness is a valid predictor of CWB. Therefore, faking behavior should have the potential to impact the correlation between CWBs and Conscientiousness in this specific case. Based on the literature just discussed, as well as our expectations for the lack of congruence between within-subjects score change and SD scores, we tested the following hypotheses:

**H5a:** Applicant faking will have a negative impact on the criterion-related validity of applicant Conscientiousness scores as predictors of self-reported CWB, such that when individuals exhibiting significant within-subjects score change are removed from the sample, validity will improve significantly.

**H5b:** Applicant faking will have a negative impact on the criterion-related validity of applicant Conscientiousness scores as predictors of self-reported CWB, such that when the applicant condition scores of individuals exhibiting significant within-subjects score change are replaced with their scores from the research administration, validity will improve significantly.

**H6a:** Partialing variance associated with SD scores from the relationship between applicant Conscientiousness and self-reported CWB will not have a significant effect on criterion-related validity.

**H6b:** Corrections for SD will be ineffective, such that when the applicant condition scores of individuals exhibiting elevated SD scale scores are replaced with their scores from the research administration, validity will not improve significantly.

As Peterson and Griffith (2006) noted, using definitions of faking that treat the behavior as a form of deception aimed at obtaining a desired outcome (e.g., Griffith & McDaniel, 2006), we may also see a link between faking and other deceptive or deviant organizational behaviors. In
a recent theoretical model of applicant faking, Goffin and Boyd (2009) noted that personality variables like integrity, in addition to a variable the authors referred to as “moral code,” may influence faking motivation. This theory is consistent with previous models of faking behavior suggesting integrity as a predictor of faking (e.g., McFarland & Ryan, 2000). If factors like low integrity and a moral code that does not deter one from engaging in deceptive behaviors are key contributors to faking motivation, then it is likely that faking should relate to other negative organizational behaviors. Although this link has not been thoroughly examined empirically, one study by Rosse, Levin, and Nowicki (1999) found that faking was related to negative work behaviors in a customer service position. An additional point worth noting that faking motivation does not necessarily translate into observed faking behavior (as it is traditionally measured). Therefore, there is a possibility that individuals who are attempting to fake may be more likely to engage in other negative behaviors, regardless of the success of their faking attempts.

H7: The amount of applicant faking will be significantly positively related to self-reported CWBs.

In the present study, we chose to examine the relationship between faking and CWBs using a difference score operationalization, as well as polynomial regression analysis (both of which are discussed in greater detail in the Method section). Polynomial regression analysis was used as a means of examining faking as a predictor of CWBs due to the concerns that have been expressed regarding the use of difference scores as predictor variables in regression analysis (e.g., Edwards, 1994) and the ability of polynomial regression to explicitly test of the constraints implicitly imposed by difference scores. In addition, an examination of the parameters of the polynomial regression model can be used to test directional hypotheses about the influence of the difference score variables on the outcome (Edwards, 1994; Edwards & Parry, 1993). Finally, the polynomial model offers a direct means of examining whether the act of faking (i.e., score change in either a positive or negative direction) predicts CWBs.

METHOD

Participants

Data for the present study were drawn from a sample of applicants to manufacturing positions in a large automotive and industrial component manufacturer. Applicants completed an online battery of assessments designed and administered by a third-party consulting firm. Descriptions of parts of this Web-based assessment system, the Select Assessment® for Manufacturing, have appeared elsewhere in the literature (c.f. Bott, O’Connell, Ramakrishnan, & Doverspike, 2007; O’Connell, Hartman, McDaniel, Grubb, & Lawrence, 2007). At the end of the battery, applicants were presented with a prompt informing them of the opportunity to take part in an independent research study. The applicants were informed that their participation in the study would result in their entry into a drawing for one of several cash prizes (up to $1,000).

In total, 3,276 applicants indicated that they were willing to be contacted about the research opportunity at a later date. This sample of individuals was then contacted by the researchers 6 weeks after completing the applicant assessments, at which time they were reinforced of the cash prize drawing and provided with a link to complete a second set of assessments. Of the 3,276
individuals who agreed to be contacted, 206 (6%) followed-through on completion of the second set of assessments. Applicant data for 10 participants were not usable, leaving a final sample of 196 for the present study. The final sample retained for analysis was 71% male and 29% female.

Given the low response rate, we chose to compare the applicant Summated Conscientiousness Scale (SCS) scores of the sample that completed the full study \( (N = 196) \) to those drawn from the sample that initially agreed to participate but did not complete the second portion of the study \( (N = 2,893) \). The two groups had similar means (88.31 and 88.64, respectively) and standard deviations (7.88 and 8.34, respectively), and an independent-samples \( t \) test confirmed that mean SCS scores from the two groups were not significantly different, \( t(2, 3087) = 1.10, \) ns. Furthermore, the \( d \)-effect size of this difference was also small \( (d = .04) \).

**Measures**

*Conscientiousness*

The SCS, a shortened version of the scale used by Griffith et al. (2007), was used as an assessment of Conscientiousness in the current study. The original scale used in the Griffith et al. (2007) study comprised 30 items, using a 7-point response scale ranging 1 (strongly disagree) to 7 (strongly agree). Griffith et al. (2007) reported an internal consistency reliability of .84, and 1-month test–retest reliability of .86. In addition, the authors reported a correlation of \( r = .76 (p < .05) \) between the SCS and the 12-item NEO-Five Factor Inventory Conscientiousness scale (Costa & McCrae, 1992). Further refinement of the scale via exploratory factor analysis resulted in the removal of the 10 items that demonstrated the lowest factor loadings on the single Conscientiousness factor. The current version of the scale comprised 20 items using a 5-point response scale ranging 1 (strongly disagree) to 5 (strongly agree).

*Applicant Faking*

Following the methodology used by Griffith et al. (2007), applicant faking was assessed via individual change scores across the applicant and research administrations of the SCS (faking = applicant condition score – research condition score). This operationalization creates a continuous faking score for each participant and has been widely used in examinations of faking behavior (Ellingson et al., 2007; McFarland & Ryan, 2000, 2006; Mueller-Hanson et al., 2003). Although the reporting and use of raw change scores as operationalizations of faking behavior is acceptable in some cases, the general use of difference scores as predictors or criteria in regression analysis has been criticized for a number of reasons (Edwards, 1993, 2001). Specifically, the use of difference scores as predictors may be problematic because they confound the effects of the original variables on which the difference is based, creating difficulties in the conceptual interpretation of significant results. In response to this and related concerns, researchers have suggested using polynomial regression as an alternative to difference scores (Edwards, 2002).

Therefore, we used a combination of polynomial regression and response surface modeling to investigate the joint influence of the Applicant and Honest SCS scores on CWBs. H7 predicts a positive relationship between the amount of applicant faking and CWBs. Because faking was operationalized as the algebraic difference between Applicant and Honest SCS scores, the
implied relationship between the difference score components (i.e., Applicant and Honest SCS scores) and CWBs is

\[ Z = b_0 + b_1 (X - Y) + e, \]

where \( X \) represents the Applicant score and \( Y \) the Honest score. This introduces a constraint that is usually not satisfied in organizational research using difference scores, namely, that the \( X \) and \( Y \) regression coefficients are equal in magnitude and opposite in sign. An unconstrained version of this equation is simply

\[ Z = b_0 + b_1 X + b_2 Y + e, \]

which allows explicit testing of the \( X \) and \( Y \) coefficients and direct interpretation of the nature of their independent relationships with CWBs.

An alternative to the algebraic difference is the use of a squared difference term \([X - Y]^2\]. The use of a squared difference may be desirable when it is hypothesized that the direction of the difference does not matter, that is, that a deviation from congruence between the difference variables in either direction (i.e., \( X > Y \) or \( Y > X \)) has a similar effect on the outcome. Although not directly hypothesized, this may be a plausible outcome in the current study, given that a mismatch between Applicant and Honest SCS scores in either direction may be related to CWBs. The unconstrained version of the regression equation reflecting the effect of the squared difference on the outcome is the quadratic model:

\[ Z = b_0 + b_1 X + b_2 Y + b_3 X^2 + b_4 XY + b_5 Y^2 + e. \]

Following guidelines outlined by Edwards and Parry (1993), we tested both the linear regression model implied by the simple algebraic difference and the quadratic model implied by the squared difference. We began by mean centering the predictors to aid in interpretation of the regression and response surface parameters. A centered score of 0.00 for the Applicant and Honest SCS corresponds to an actual score of 88.31 and 86.80, respectively.

Next, we estimated regression equations for the unconstrained linear and quadratic models. These can be considered competing models with different levels of complexity and parsimony. Therefore, we used hierarchical regression to determine whether the quadratic model accounts for additional variance in CWBs above and beyond the linear model, which it subsumes. Table 4 shows that the quadratic model accounted for significant variance in the outcomes above and beyond the linear model. Therefore, the parameters of the quadratic model were used to describe the joint relationship of Applicant and Honest SCS scores on CWBs. In addition, in the final step, a response surface was estimated to depict this relationship, and the regression parameters were used as input into equations for estimating key features of the response surface.

Finally, to identify individuals who exhibited considerable score change (and are subsequently referred to as “fakers”), we also used the confidence interval methodology established by Griffith et al. (2007; a similar methodology was also employed by Hogan et al., 2007). The goal of this methodology is to identify individuals whose score change exceeds the boundaries that may be expected when random measurement error is taken into account. The result is the establishment
of what could essentially be referred to as a “faking cut score.” Specifically, the standard error of measurement (SEM) for the SCS (derived from the reliability and standard deviations from the research administration) is multiplied by 1.96, creating a 95% confidence interval around the research context score. If an individual’s score in the applicant condition exceeds the upper bound of this confidence interval, he or she is flagged as a “faker.” Within the current sample, the SEM was equal to 2.66, making the faking cutoff value equal to 5.22 points on the SCS.

**Social Desirability**

A short form of the Marlowe-Crowne (Crowne & Marlowe, 1960) SD scale was used in the present study. The Marlowe-Crowne is a widely used assessment of SD response bias. A 10-item form created by Strahan and Gerbasi (1972) was used in lieu of the full 33-item version of the assessment. A study by Fischer and Fick (1993) found that this particular short form correlated at $r = .96$ with the full 33-item scale and demonstrated adequate internal consistency ($\alpha = .88$).

**CWBs**

CWBs were assessed using a modified 38-item self-report measure of CWB developed by Bennett and Robinson (2000) and Robinson and O’Leary-Kelley (1998). Evans (2005) combined items from workplace deviance assessments developed by Bennett and Robinson and by Robinson and O’Leary-Kelly with a focus on assessing day-to-day CWBs (e.g., deliberately wasting time) rather than extreme CWBs (e.g., workplace violence). Evans reported an internal consistency reliability estimate of $\alpha = .94$. Scale items were summed to create a CWB scale score.

**Procedure**

All participants completed the SCS and Marlowe-Crowne short form during the online screening process carried out by the independent consulting firm. The SCS and Marlowe-Crowne scales were embedded in a battery of other personality and ability assessments used to make initial selection decisions. At the end of the screening process, all applicants were presented with a visual prompt indicating that they had the opportunity to participate in an independent research study being conducted by a university that was unaffiliated with the hiring organization. Applicants were asked to indicate whether they wished to be contacted by the researchers at a later date.

All individuals providing valid e-mail addresses were contacted by the researchers 6 weeks after completing the applicant screening process. The initial e-mail contact made by the researchers reminded the individuals that they had agreed to be contacted for an independent research study and provided participants with a link to a secure Web site operated by the university, which contained the second set of assessments. The researchers were unaware of the participants’ status in the hiring process at this point. Given the large scale of the screening process, in many cases, the respondents were still awaiting information regarding their standing in the hiring process.
After consenting to take part in the study, participants were taken to an instruction page which presented specific instructions to respond honestly to the questionnaire items. The instructional set used was as follows:

For the following set of questionnaires, you are asked to answer the questions in such a way that would most resemble yourself. Your goal is to answer as HONESTLY as possible. It is very important that you answer as honestly as possible, **even if the description would be considered unflattering**.

... Please take your time and read each item carefully. Remember that this information is **completely confidential, and that your current/prospective employer will not have access to your questionnaire**. The only individuals that will be allowed access to the questionnaire are the researchers involved.

After reading the instructional set, participants began the online assessment, which contained the items from the SCS and the CWB scales. Participants provided demographic information at the end of the online questionnaire.

**RESULTS**

Descriptive statistics and estimates of internal consistency are available in Table 1, whereas intercorrelations among the variables assessed in the study appear in Table 2. All scales demonstrated adequate levels of internal consistency. In addition, the difference scores used as a measurement of faking behavior also demonstrated adequate reliability ($r_{dd} = .75$).

**The Prevalence of Faking Behavior**

H1 proposed that significant mean-level differences in Conscientiousness scores would be observed between the applicant and research administrations of the assessment. A repeated measures analysis of variance provided support for this assertion, indicating that at the group level of analysis applicant scores ($M = 88.31$) were significantly higher than scores collected in the research context ($M = 86.81$), $F(1, 195) = 8.61$, $p = .004$. Further examination of Cohen’s $d$ effect size suggested that although this effect was statistically significant, it was not large ($d = .18$).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Descriptive Statistics and Reliability Estimates for Study Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min.</td>
</tr>
<tr>
<td>1. Applicant Conscientiousness</td>
<td>69.00</td>
</tr>
<tr>
<td>2. Honest Conscientiousness</td>
<td>60.00</td>
</tr>
<tr>
<td>3. Amount of Faking</td>
<td>−19.00</td>
</tr>
<tr>
<td>4. Social Desirability</td>
<td>27.00</td>
</tr>
<tr>
<td>5. Counterproductive Work Behaviors</td>
<td>38.00</td>
</tr>
</tbody>
</table>

Note. $N = 196$.

*Reliability for the amount of faking variable was estimated using the formula for the reliability of difference scores provided by Crocker and Algina (1986).
H2 stated that a significant portion of the sample would be identified as having faked the Conscientiousness assessment, using Griffith et al.'s (2007) faking cutoff methodology. Overall, 24% of the sample met the requisite criterion (i.e., applicant score exceeded honest score by more than 5.22 points) for being identified as having faked the personality assessment. A chi-square test suggested that this percentage represented a significantly higher portion of the sample than would be expected solely due to measurement error ($\chi^2 = 148.6, p < .001$). The results of this analysis provide support for the second hypothesis.

**SD as an Indicator of Faking**

H3 and H4 pertained to the effectiveness of the SD scale as an indicator of faking or within-subjects score change across the applicant and research administrations. As an examination of the correlation matrix (Table 2) reveals, SD scale scores were not significantly related to applicant faking ($r = .14, ns$). This finding supports H3.

H4 stated that attempts at identifying respondents suspected of faking on the Conscientiousness assessment using a measure of SD would be statistically independent of identifications based on significant within-subjects score change. A Pearson’s chi-square test of independence was employed to examine the independence of the two classifications. Given that no suggested “cut-score” for flagging respondents on the SD scale used in the present study exists, we tested this hypothesis using the ranges of SD scores used in Schmitt and Oswald’s (2006) simulation. The authors examined the impact of corrections for SD scores across conditions in which the top 5%, top 15%, and top 30% of scorers on the SD measure were suspected of faking. Dummy coded variables representing whether each individual respondent would have been flagged were created for each simulated cutoff.

Across all three SD cut scores, the identification of faking using SD scores was statistically independent of identification based on Griffith et al.’s (2007) confidence interval methodology. The situation in which the top 30% of scorers on the SD scale were flagged as fakers represents the most direct test of this hypothesis, given that the comparison involves similar absolute percentages of individuals being flagged. Even in this situation, the chi-square test of independence was nonsignificant ($\chi^2 = .88, ns$). Table 3 presents the full chi-square statistics as well as a display of hit–miss ratios for the SD scale. At the top 5% SD cutoff, none (0%) of the individuals

---

1In this analysis, independence is the null hypothesis, with statistical significance indicating the presence of a relationship between the two variables.
flagged based on their SD scores exhibited significant within-subjects score change. This hit rate improved at the 15% SD cutoff (24%), before dropping off again at the 30% SD cutoff (19%). Overall, the tests of H3 and H4 suggest that the SD scale was ineffective at identifying individuals who exhibited significant score elevation in the applicant context.

**Faking and CWBs**

H5 (a and b) and H6 (a and b) were aimed at testing the effects of applicant faking and SD on the criterion-related validity of Conscientiousness as a predictor of self-reported CWBs. Specifically, H5 stated that the removal of individuals suspected of faking on the Conscientiousness measure (using Griffith et al.’s methodology) would result in a significant improvement in criterion-related validity. As Table 2 indicates, the criterion-related validity of applicant Conscientiousness scores as a predictor of CWBs for the full sample was low ($r = -0.05$, ns). However, additional analyses suggest that this low level of criterion-related validity was due, at least in part, to faking.

First, a subgroup correlation analysis for the 76% of the sample that were not flagged for suspicion of faking ($n = 149$) resulted in a considerable improvement in criterion-related validity ($r = -0.18$, $p < .05$). The results of a $t$ test for dependent correlations indicated that this improvement was statistically significant ($t = 3.32$, $p < .01$), which supports H5a. Next, we examined the criterion-related validity of the applicant Conscientiousness assessment by replacing the applicant scores of those individuals identified as fakers with their scores from the honest condition. This resulted in a more substantial gain in validity ($r = -0.22$, $p < .01$). Once again, this gain in validity was found to be statistically significant ($t = 3.82$, $p < .01$). This finding lends support to H5b.

In comparison to the findings regarding the impact of faking on criterion-related validity evidenced in H5, H6a suggested that partialing variance associated with SD scores from the correlation between applicant Conscientiousness scores and CWB would not significantly improve validity. This technique has been relied upon by notable studies that have been cited as evidence that faking is not a threat to validity (e.g., Ones et al., 1996). After controlling for the effects of SD scores, the correlation between applicant Conscientiousness and counterproductive behavior remained nonsignificant ($r = .10$, ns) and actually moved in the direction of being degraded. This finding supports the assertion made in H6a. In addition, to provide a parallel test of H5b, we examined the validity of Conscientiousness when the applicant condition scores of the top 5%, 15%, and 30% of scorers on the SD scale were replaced with their honest condition scores (H6b). For the 5% SD score cutoff, the criterion-related validity of Conscientiousness
remained at $r = -.05$, $ns$. Marginal improvements were found for the 15% ($r = -.10$, $ns$) and 30% ($r = -.10$, $ns$) SD cutoffs using the replacement procedures, though in neither case was the validity coefficient significant.

Finally, H7 stated that the amount of faking variable would be significantly positively related to self-reported CWBs. As an examination of Table 2 indicates, this was in fact the case ($r = .22$, $p < .01$). However, as previously stated, using change scores to measure faking may be problematic, and these results could be due to a statistical artifact. Therefore, polynomial regression and response surface methodology were employed as complimentary tests of H7, allowing for an examination of the prediction of CWBs as a function of faking. Table 4 shows the parameters of the unconstrained linear and quadratic regression equations used to test H7 within the polynomial regression framework previously described. The quadratic model accounted for significant variance in CWBs above and beyond the linear model. This suggests that deviations between the Honest and Applicant SCS scores in either direction lead to high levels of CWBs. The coefficients for the Honest SCS linear term, the Honest SCS quadratic term, and the Applicant × Honest interaction term were found to be significant.

Our inspection of the response surface corresponding to this model revealed a convex (i.e., valley-shaped) shape, with the edges of the surface curving upwards away from the $x = y$ line. Key features of the response surface are the stationary point (i.e., the highest or lowest point on the surface at which the slope is 0 in all directions) and the principal axes (perpendicular lines that intersect at the stationary point). The stationary point was found to be near the center of the surface ($x = -0.89$, $y = -0.30$). The first principal axis was found to be $Y = -1.40 + -0.93X$; the upward curvature is greatest along this line. The second principal axis, along which upward curvature is least, was determined to be $Y = 0.08 + 0.32X$.

The regression results suggest that any deviation from congruence (i.e., where applicant scores are equal to honest scores) can be problematic, with CWBs increasing as scores become incongruent. However, there is a significantly greater impact on CWBs when Applicant SCS scores are higher than Honest SCS scores (an indication of positive faking). Overall, the results of these analyses suggests that individuals who engage in faking to a greater degree (even unsuccessfully faking in the wrong direction) also tend to report engaging in more CWBs, thereby supporting H7.

### TABLE 4

Polynomial Regression Analysis Results

<table>
<thead>
<tr>
<th>Model</th>
<th>$\beta$</th>
<th>sig.</th>
<th>Model $R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta F$</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Linear model</td>
<td>—</td>
<td>—</td>
<td>0.06</td>
<td>0.06</td>
<td>6.62</td>
<td>**</td>
</tr>
<tr>
<td>Applicant SCS</td>
<td>0.15</td>
<td>$ns$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Honest SCS</td>
<td>$-0.32$</td>
<td>*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Model 2: Quadratic model</td>
<td>—</td>
<td>—</td>
<td>0.11</td>
<td>0.05</td>
<td>3.30</td>
<td>**</td>
</tr>
<tr>
<td>Applicant SCS</td>
<td>0.12</td>
<td>$ns$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Honest SCS</td>
<td>$-0.27$</td>
<td>**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Applicant Squared</td>
<td>0.09</td>
<td>$ns$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Honest Squared</td>
<td>0.19</td>
<td>**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Applicant × Honest</td>
<td>$-0.32$</td>
<td>**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. SCS = Summated Conscientiousness Scale.

* $p < .05$. ** $p < .01$. 
In contrast, SD scale scores were actually negatively related to counterproductive work behavior ($r = -0.18$, $p < 0.05$). This finding is explored in greater detail in the Discussion section.

**DISCUSSION**

The current study used a within-subjects design to examine the prevalence of faking behavior in a sample of real job applicants, the validity of SD scales as indicators of faking, and faking as both a correlate of and a threat to the prediction of CWBs. According to both Ellingson et al. (2007) and Hogan et al. (2007), within-subjects designs with real applicant data represent the most direct test of faking behavior currently available to researchers. We concur with the aforementioned authors in that we believe the use of a similar design enhances the present study’s contribution to the extant body of knowledge in the realm of applicant faking behavior. The results of this study, although supporting earlier findings by Arthur et al. (2010) and Griffith et al. (2007), were inconsistent with the findings of Ellingson et al. (2007) and Hogan et al., who both reported little evidence for the occurrence of faking in applicant samples. In contrast, the results of the present investigation suggest that a sizeable portion of job applicants engage in faking behavior.

**Summary of Findings**

Twenty-four percent of participants displayed significantly elevated personality scores in the applicant condition, a finding that is relatively consistent with Donovan et al.’s (2003) findings in a study of self-reported faking behavior, and Griffith et al.’s (2007) study of faking in applicants to a temporary employment agency. Although this result is at odds with the investigations by Ellingson et al. (2007) and Hogan et al. (2007), we believe that a variety of methodological factors may have resulted in the lack of faking behavior observed in these studies. Specifically, the wide variations in retest intervals observed in the Ellingson et al. (2007) study, and the applicant-applicant retest setting used by Hogan et al. may have played a role in the variation in findings. The data reported herein, gathered across applicant and research contexts with a consistent retest interval, suggest that faking does occur in applicant samples.

Perhaps the most revealing findings gleaned from this investigation were our results pertaining to the use of SD scales and the prediction of CWBs. Our findings suggest that SD scales are poor indicators of faking when the behavior is directly measured via score changes across applicant and research contexts. The SD measure used in the present study was not significantly related to the amount of score change. Given that SD scores are typically used as a proxy variable for faking behavior in both research (e.g., Ones et al., 1996) and practice (e.g., Goffin & Christiansen, 2003), this finding is cause for concern. According to Hunter and Schmidt (2004), the quality of a proxy variable can be assessed by multiplying the correlation between the proxy variable (i.e., SD) and the behavior of interest (i.e., faking) by the reliability coefficient for the proxy variable. In the current study, this results in an estimate of SD as a proxy variable of .10, which is disappointingly low.

In addition, our findings indicated that the SD scale was not readily able to correctly identify individuals who had exhibited significant levels of score change across contexts. In fact, in the best-case scenario (within the range of possibilities examined), the SD scale was only able to
correctly identify individuals exhibiting significant within-subjects score change 21% of the time. Perhaps more important, the false positive and false negative rates for the SD scale were also high. False positive rates (i.e., individuals flagged by the SD scale who did not exhibit significant within-subjects score change) ranged from 81% (at the 30% SD cutoff) to 100% (at the 5% SD cutoff). In addition, even when the 30% SD cutoff was applied (therefore identifying roughly the same portion of the sample as the Griffith et al., 2007, methodology), 79% of the individuals identified as exhibiting significant score change were not flagged by the SD scale. Such high false-positive rates could result in the unfair penalization of individuals who score high on the SD scale, whereas the high false-negative rates suggest that many of the individuals who actually elevate their scores would not be identified. Taken together, our findings regarding the accuracy of SD scales suggest that many assertions regarding faking behavior that have been based on the use of these scales as proxy variables (e.g., Barrick & Mount, 1996; Ones et al., 1996) may be unfounded.

Although we discuss the lack of correspondence between the SD scores and within-subject score change as evidence that SD scales may be problematic, we believe it is worthwhile to note potential limitations of the applicant-research score change measure as well.² Specifically, it is plausible that some individuals attempted to fake not only in the applicant context but also in the research context (thereby making the score change estimate and classification an underestimate of the true prevalence of faking behavior) in order to present a consistent image, or out of skepticism regarding the confidentiality of their responses. In addition, the score change measure does not adequately capture faking attempts that do not result in positive score increases in the applicant condition. For example, it is possible for an individual to be attempting to fake but doing so in a positive direction for some items and a negative direction for others. Recent work by Kuncel and Tellegen (2009) indicates that respondents’ faking attempts are often complex and may not always result in linear score increases as is typically predicted. In this case, the score change measure would likely indicate small to nonexistent amounts of faking behavior, whereas the confidence interval classification would also likely fail to identify this individual as a “faker.” Rather than making a definitive statement regarding SD as a predictor of faking, we simply observed and reported the incongruence between those SD scores and within-subjects score change. We believe that the findings regarding differences in the identification of fakers using the score change and SD methods offer a useful addition to the existing discourse on this topic (e.g., Burns & Christiansen, 2006; Griffith & Peterson, 2008).

Our findings regarding faking and CWBs are also noteworthy. First, our data suggest that faking has the potential to attenuate correlations between personality constructs and other relevant organizational behavior variables. Thus, our findings were consistent with recent research (e.g., Komar et al., 2008; Peterson et al., 2009) examining faking and attenuation. The Conscientiousness measure used in the current study was not significantly related to CWBs for the full sample. However, when the applicant scores of individuals identified as having faked the personality assessment were replaced with scores from the honest condition, the Conscientiousness measure demonstrated a considerably stronger relationship with the self-report CWB scale ($r = -0.22, p < .01$). Finally, this same effect on the Conscientiousness–CWB

---

²We thank an anonymous reviewer for this suggestion.
relationship was not observed when correcting for SD scores, therefore adding to the evidence for a reevaluation of research findings based on assessing faking using SD scales.

In addition to the negative effect of faking on the relationship between Conscientiousness and CWBs, our findings also suggest that in some cases faking may be related to other negative workplace behaviors. We found a significant positive correlation between faking behavior and self-reported CWBs in addition to further supporting this finding using polynomial regression analysis. The polynomial regression results suggested that faking in the positive and negative directions was linked to increases in self-reported CWBs. This finding may indicate that the act of attempting to fake, regardless of whether this attempt results in score increases, may be related to negative organizational outcomes. Our findings regarding CWBs support the call for broadening considerations of the impact of faking behavior (Peterson & Griffith, 2006). Furthermore, scores on the SD scale were related to CWB in the opposite direction (i.e., high SD scores were associated with lower self-reported CWB). This finding could be viewed as support of conceptualizations of SD as a substantive trait (e.g., Smith & Ellingson, 2002), such that some individuals who score high on measures of SD actually tend to engage in more socially acceptable behaviors (therefore engaging in low levels of counterproductive behavior).

Implications

Although we have made reference to the position of our findings in the broader scope of the extant literature pertaining to applicant faking, there are several key implications for research and practice that can be drawn from our findings. First, although we believe that the use of within-subjects investigations allows for the most detailed examination of faking at the individual level, the widely varying results of recent research employing these designs suggest the need for more research efforts. Researchers must examine the nature of specific testing situations (e.g., research setting, retest settings, personal development settings) to fully explicate their influences on the occurrence of faking behavior. Recent discussions of faking from the interactionist perspective (e.g., trait activation theory; Tett et al., 2006) would suggest that subtle differences in situational cues can exert varying influences on responses to personality items.

In addition, we believe that our findings should raise a great deal of concern over the use of SD measures. Given the high-stakes nature of employment testing, using measures that may not capture actual faking behavior may represent an additional threat to the validity of personnel decisions rather than a guard against faking. Considerable support for the use of these assessments, using real applicants and validation against other measures of faking, appears to be lacking.

A final implication of our findings is that faking may represent not only a threat to validity but also a correlate of other forms of deceptive behavior. The link between faking behavior and self-reported CWBs evidenced in the present study suggests a need for more detailed investigations of the degree to which faking behavior relates to other deceptive behaviors. It is possible that faking and CWBs share a common determinant such as low self-control, which has been linked to a variety of deviant behaviors (Gottfredson & Hirschi, 1990). If individuals who engage in faking on personality assessments are able to do so to a degree that allows them to gain entry into organizations, there appears to be a potential for them to engage in additional acts of organizational deviance. Although this certainly has implications for organizations that use personality
assessments, we believe it should also spur more research that focuses on the actual on-the-job behavior of individuals who engage in faking.

Limitations

Several limitations to the current study are worth noting. First, the response rate was less than desirable. This is one of several potential methodological limitations inherent in any within-subjects investigation of applicant faking in an applied setting. In our case, the low response rate was a limitation of attempts at gathering honest responses in the research context. The difficulties associated with gathering data in this manner were also evident in the study by Arthur et al. (2010), who reported a response rate of roughly 3% for a similar design. Although the possibility for unmeasured differences (e.g., economic need/income differences, etc.) between those who completed the second phase of the study and those who did not exists, the fact that applicant SCS scores were so similar across the two groups should assuage some concern over the potential for sampling error associated with the low response rate.

An additional limitation was the use of a self-report assessment as a criterion variable. However, CWBs represent a challenging variable for researchers to assess in an objective fashion. As Berry et al. (2007) pointed out, even supervisor and peer reports of CWB are subject to the influences of halo bias and low detection rates. In addition, Berry et al. noted that the self-report method is the most common form of CWB measurement. Although we believe that our findings represent valid inferences concerning the relationship between faking and CWBs, the extension of this finding to measures of CWB that do not rely on self-report is certainly needed to bolster its generalizability.

Finally, the current study’s reliance on single measures of social desirability, Conscientiousness, and a single criterion measure may also limit the generalizability of our findings. Most notably, given that a primary focus of the study was to distinguish within-subjects measures of faking behavior from measures of socially desirable responding, our conclusions should be considered against the fact that only one measure of SDR was examined. As Paulhus (2002) noted, measures of SDR can vary considerably in their development and focus.

Conclusions

The results of this study shed considerable light on the nature of faking behavior in applicant samples. In our assessment, it appears that faking is difficult to capture using SD scales. In addition, faking appears to have the potential to affect criterion-related validity while also being related to negative workplace behaviors (CWBs). These findings should spur continued efforts aimed at developing effective operational methods of identifying individuals who fake as well as understanding the range of implications that faking behavior may have for organizations.

ACKNOWLEDGMENTS

Mitchell H. Peterson is now at Globe University/Minnesota School of Business. We thank Paul Sackett, Michael McDaniel, Michael Zickar, Patrick Converse, and Fred Oswald for their comments and suggestions on an earlier draft of this article.
REFERENCES


