2-2005

Situational judgment tests and their predictiveness of college students' success: The influence of faking

Helga PEETERS
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg
DOI: https://doi.org/10.1177/0013164404268672

Published in Educational and Psychological
Measurement, 2005 February, Volume 65,
Issue 1, Pages 70-89
http://doi.org/10.1177/0013164404268672

SITUATIONAL JUDGMENT TESTS AND THEIR PREDICTIVENESS OF COLLEGE STUDENTS' SUCCESS: THE INFLUENCE OF FAKING

HELGA PEETERS AND FILIP LIEVENS
Ghent University, Belgim

There is increasing interest in using situational judgment tests (SJTs) to supplement traditional student admission procedures. An important unexplored issue is whether students can intentionally distort or fake their responses on SJTs. This study examined the fakability of an SJT of college students' performance. Two hundred ninety-three psychology students completed a cognitive test, a personality measure, and an SJT. Only for the SJT, the students were assigned to either an honest or a fake condition. The scores of students in the fake condition were significantly higher than those of students in the honest condition ($d = .89$). Furthermore, faking had a negative effect on the criterion-related validity (there was a significant drop from $r = .33$ to $r = .09$) and the incremental validity of the SJT over cognitive ability and personality. These results are discussed in terms of the use of SJTs in high-stakes testing programs.

Keywords: *situational judgment; faking; response distortion; academic success; validity*

In personnel selection, situational judgment tests (SJTs) have emerged as an important and useful complement to traditional cognitively oriented tests (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Motowidlo, Dunnette, & Carter, 1990). SJTs present applicants with verbal descriptions of hypothetical job-related scenarios and ask them to identify an

We would like to acknowledge Mike Zickar for his valuable suggestions on an earlier version of this article and Morrie Mullins for making the SJT available. Portions of this article were presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology (Orlando, FL). Correspondence concerning this article should be addressed to Helga Peeters, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium; e-mail: helga.peeters@ugent.be.

appropriate response from a list of alternatives (Motowidlo et al., 1990; Motowidlo, Hanson, & Crafts, 1997; Weekley & Jones, 1999). SJTs are becoming increasingly popular in personnel selection for various reasons. First, large-scale studies have shown that SJTs have significant criterion-related validity (McDaniel et al., 2001) and possess incremental validity over and above cognitive ability and personality tests (Chan & Schmitt, 2002; Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001). Second, people respond enthusiastically to SJTs because they perceive SJTs to be related to the target jobs for which they are applying (Kluger & Rothstein, 1993; Ployhart & Ryan, 1998). A third important advantage is that SJTs show less adverse impact on minorities than traditional cognitive ability tests (Clevenger et al., 2001), although it has been argued that this may be the result of the lower reliability of SJTs (Chan & Schmitt, 1997; Schmitt, Clause, & Pulakos, 1996).

Given these advantages, there exists growing interest in using SJTs to supplement traditional admission techniques in high-stakes testing programs (Lievens, Buyse, & Sackett, in press; Lievens & Coetsier, 2002; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Sackett, 2002). Traditionally, the selection procedure for admission to undergraduate and graduate university programs has been based on prior academic achievement (grade point average [GPA]) and cognitively oriented predictors. The latter is exemplified by the use of the Scholastic Aptitude Test (SAT), the American College Test (ACT), and the Graduate Record Examinations (GRE) General Test as admission instruments. Although large-scale meta-analyses (Hezlett et al., 2001; Kuncel, Hezlett, & Ones, 2001) have provided evidence for the predictive validity of these tests, over the years, concerns have been raised that academic admissions are based too much on cognitively oriented measures for deciding which students may gain access to higher education (Atkinson, 2001; Oswald et al., 2004; Sternberg, Wagner, Williams, & Horvath, 1995). This overuse of cognitively oriented predictor measures in high-stakes testing programs might be problematic because these measures have been shown to have an adverse impact on minorities (Hough, Oswald, & Ployhart, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt, Sackett, & Ellingson, 2002). In addition, because of the heavy emphasis on cognitive ability, a wider range of important skills and traits (e.g., interpersonal skills, leadership, etc.) is ignored, calling for alternative forms of student selection and admission.

Initial evidence for the use of SJTs in predicting student performance is encouraging. For example, Oswald et al. (2004) constructed an SJT for predicting student performance and established its predictive validity in terms of predicting students' self-reports of various dimensions of student performance. In addition, they found that nearly all SJT scales had incremental validity over personality and cognitive measures when predicting college students' performance. In a similar vein, Lievens et al. (in press) found that an

interpersonally oriented SJT became more valid through the years and was especially valid for students' scores on interpersonal courses.

Although initial evidence for the use of SJTs in a student admission context is encouraging, it is clear that other issues still need to be addressed before SJTs can become a part of high-stakes testing programs. One of these unexplored issues is whether students might improve their performance on SJTs by faking good. If SJTs are not prone to faking effects, then an important threat to the operational use of SJTs in high-stakes testing is removed. This might provide decision makers with additional evidence in favor of the adoption of SJTs in these testing programs. Conversely, if faking seriously affects SJTs' performance, decision makers might first turn their attention to strategies to reduce faking on SJTs.

Therefore, this study investigated whether people are able to fake good and raise their scores on an SJT of college students' performance (see Bess & Mullins, 2002) if instructed to do so. In particular, the influence of faking is investigated in three ways: by comparing (a) mean scores, (b) criterion-related validity, and (c) incremental validity across honest and fake conditions. This study also explored the relationships between faking on an SJT and the constructs being measured.

Study Background

Faking in Personnel Selection

Research on faking, or intentional response distortion, has a long tradition in personnel selection (Ones, Viswesvaran, & Reiss, 1996). Faking a selection measure can be defined as an individual's conscious distortion of responses to score favorably (e.g., Dwight, 1999; McFarland & Ryan, 2000). Faking has especially been studied in non-cognitive-oriented measures, such as personality tests, biodata inventories, and integrity tests (Alliger & Dwight, 2000; Becker & Colquitt, 1992; Dalen, Stanton, & Roberts, 2001; Graham, McDaniel, Douglas, & Snell, 2002; Kluger & Collela, 1993; Martin, Bowen, & Hunt, 2002; McFarland, Ryan, & Aleksander, 2002; Ones & Viswesvaran, 1998).

In the faking literature, several research streams can be distinguished (Viswesvaran & Ones, 1999). One stream of faking research (e.g., Cowles, Darling, & Skanes, 1992; Martin et al., 2002; Zickar & Robie, 1999) involves experimental studies comparing responses obtained under different instruction sets (instructions to be honest or to fake good). Generally, this body of research is conducted in laboratory settings and aims to determine the maximal limits of faking capability. Comparing responses obtained under instructions to be honest and instructions to fake good provides the maximum limits of the extent to which test scores can be changed by faking to make a positive

impression (Viswesvaran & Ones, 1999). These experimental studies use either a within-subjects or a between-subjects design (Furnham, 1986), with both having their merits and limitations (Cook & Campbell, 1979; Edens & Arthur, 2000). There is relative consensus in the literature that applicants can distort their answers on noncognitive measures and improve their scores through faking if instructed to do so. In fact, Ones, Viswesvaran, and Korbin's (1995) meta-analysis reported that faking can increase scores by nearly one half standard deviation. However, this does not demonstrate that applicants actually fake in organizational settings or on real-life applications (e.g., Hough & Schneider, 1996; Rosse, Stechner, Levin, & Miller, 1998).

A second stream of research (e.g., Smith & Ellingson, 2002) involves field studies, comparing responses from various groups (e.g., students, applicants, incumbents). These studies aim to determine the typical and operational levels of faking in real-world settings. The meta-analysis of Edens and Arthur (2000) confirmed that real-life motivational distortion resulted in smaller effect sizes ($d = .30$) than instructionally induced faking in laboratory studies ($d = .73$), suggesting that laboratory findings may be a worst-case scenario in comparison with faking in actual selection situations (e.g., Griffith, Frei, Snell, Hamill, & Wheeler, 1997; Rosse et al., 1998; Weiner & Gibson, 2000). Moreover, Edens and Arthur analyzed a subset of real-life motivational distortion (viz., job applicants compared with incumbents), providing an estimate of the magnitude of distortion that occurs in a typical personnel selection setting, and found a quite small effect size ($d = .17$).

The debate over the extent to which applicants actually engage in faking in employment settings has yet to be resolved. Some studies have reported that applicants do not fake personality tests, and even if they do, it does not negatively affect their validity (Abrahams, Neumann, & Githens, 1971; Cunningham, Wong, & Barbee, 1994; Ellingson, Smith, & Sackett, 2001; Hough, 1998; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; McCrae & Costa, 1983; Ones & Viswesvaran, 1998; Ones et al., 1996). Conversely, other studies have found that faking occurs in selection settings and attenuates the criterion-related validity of personality tests (Anderson, Warner, & Spencer, 1984; Douglas, McDaniel, & Snell, 1996; Dunnette, McCartney, Carlson, & Kirchner, 1962; Kluger, Reilly, & Russell, 1991; Rosse et al., 1998; Schmit & Ryan, 1992; Worthington & Schlottmann, 1986; Zickar, 1997). For example, Schmit and Ryan (1992) found that the validity of a test for predicting GPA was lower for students who received an incentive to misrepresent themselves. Furthermore, researchers have recently turned their attention to practical outcomes associated with faking and discovered that faking can have a significant effect on who is hired (Rosse et al., 1998; Weiner & Gibson, 2000).

Instead of comparing the responses of various groups, the operational level of faking in applicant settings is often defined by scores on social desir-

ability measures, assuming that those with high social desirability scores have faked, whereas low scorers are assumed to have responded honestly (Paulhus, 1986, 1991). However, the most important criticism is that social desirability is not, as commonly assumed, synonymous with faking. According to several researchers (e.g., McCrae & Costa, 1983; Nicholson & Hogan, 1990), social desirability scales measure important individual differences in personality constructs, and controlling for social desirability actually reduces validity. Additionally, when using such scales, caution must be taken to avoid unduly penalizing honest respondents because these scales treat extreme respondents with high scores as if they were faking.

A third stream of faking research focuses on better understanding the essence of applicants' faking behavior (McFarland & Ryan, 2000; Snell, Sydell, & Lueke, 1999). Therefore, models of faking behavior are developed that specify the antecedents and moderators (e.g., situational characteristics, personality characteristics) of faking behavior. In a similar vein, other researchers (Zickar & Robie, 1999) use item response theory to model the processes underlying faking.

Faking SJTs

As opposed to the extensive research on the fakability of personality, biodata, and integrity tests, no published studies have examined whether people can intentionally distort their responses on SJTs. Granted, a limited number of conference presentations (Haas & McDaniel, 1999; Juraska & Drasgow, 2001; Nguyen, McDaniel, & Biderman, 2002) have investigated whether individuals are able to fake SJTs when instructed to do so. To date, these studies' results have been mixed, with some showing that SJTs can be faked (Haas & McDaniel, 1999; Nguyen et al., 2002) and others (Juraska & Drasgow, 2001) drawing the opposite conclusion.

In Haas and McDaniel's (1999) study, participants completed the Work Problem Survey under both honest and fake instructions. The results revealed that fakers improved their scores by one half standard deviation. Conversely, Juraska and Drasgow (2001) concluded that SJTs were not fakable. In this study, participants were randomly assigned to an honest or fake condition and they completed either a multimedia SJT (the Conflict Resolution Skills Assessment) or a paper-and-pencil personality measure (the Assessment of Background and Life Experiences). Contrary to the mean personality scores, the mean SJT scores did not differ significantly between the honest and fake conditions, suggesting that the SJT was resistant to faking. Nguyen et al. (2002) posited that these mixed results regarding the fakability of SJTs might result from the types of response instructions given (see also Ployhart & Ehrhart, 2001). Nguyen et al. found that SJTs were more resistant to faking when respondents were asked to indicate the best and the worst responses

(i.e., the knowledge condition) than when they were asked to indicate the most and the least likely responses (i.e., the behavioral tendency condition). They also found that SJTs were more cognitively saturated in the knowledge condition.

Although the results of these prior studies are interesting, they share three important limitations. First, these studies did not examine whether faking affects the validity of SJTs. They examined only whether faking affects SJT performance. Clearly, a more important research question is whether faking affects the criterion-related validity of SJTs. Therefore, our study extended prior studies by examining the influence of faking in a more comprehensive way. We compared the mean scores on an SJT across two groups (honest and fake). Additionally, we investigated to what extent faking on the SJT affected its criterion-related validity.

Second, prior studies have not examined whether faking affects the incremental validity of SJTs. As already noted, recent research has demonstrated that SJTs have incremental validity over and above cognitive ability and personality tests (Chan & Schmitt, 2002; Clevenger et al., 2001; Oswald et al., 2004). However, it might be that when people are instructed to fake an SJT, the SJT no longer explains additional variance over and above more commonly used predictors, such as personality and cognitive ability. Even in the broader faking literature, we are not aware of any studies that have used incremental validity as a dependent variable. This lack of research is surprising considering that the incremental validity of a selection procedure is of paramount practical and theoretical importance (Schmidt & Hunter, 1998). Therefore, our study investigated whether faking on an SJT affects its incremental validity.

A third shortcoming of prior research is that the fakability of SJTs has been examined without inspecting the correlations of SJTs with other constructs. Because an SJT is best viewed as a measurement method designed to assess a variety of constructs (Chan & Schmitt, 1997; McDaniel et al., 2001; Weekley & Jones, 1999), the constructs measured by one particular SJT might not be the same constructs measured by another SJT, or at least the magnitude of the correlation with the constructs might be variable across SJTs. Therefore, an alternative explanation for the aforementioned mixed results might stem from differences in the constructs being captured by the various SJTs. Specifically, the fakability of SJTs might depend on their correlation with one particular construct, namely, cognitive ability (see Nguyen & McDaniel, 2001). In their meta-analysis, McDaniel et al. (2001) estimated the average correlation between SJTs and cognitively oriented constructs to be .46. The 10% to 90% confidence interval ranged from .17 to .75, indicating that there was substantial variability around the mean correlation and that some SJTs might be more *g*-loaded than others. Because faking on cognitively oriented measures is difficult (one either knows the correct answer or

does not), *g*-loaded SJTs might be more resistant to faking. Conversely, people might be better able to fake good and improve their performance on SJTs when the SJTs do not show substantive correlations with cognitive ability. Therefore, an exploratory part of this study was to examine the relationships between faking and the constructs measured by SJTs at the overall score level as well as at the individual item level.

Method

Sample and Procedure

A total of 293 second-year psychology students from a large Belgian university participated in the study as part of introductory courses about psychological testing and assessment. The sample was predominantly female (81%). Participants were between 19 and 23 years old ($M = 20.6$ years, $SD = 2.32$ years).

Research assistants administered a series of psychological tests in the following order: a cognitive ability test, a personality measure, and an SJT. For the SJT only, the participants were randomly assigned to an honest ($n = 138$) or a fake ($n = 153$) condition. The specific instructions in both conditions were adapted from previous studies (e.g., McFarland & Ryan, 2000; Nguyen et al., 2002). In the honest condition, participants were asked to

> decide which response is the answer choice that best describes what you would do in that situation. Please answer the questions as honestly as possible. We are not asking to put yourself in a favorable manner, but rather how you would really handle the situation.

In the fake condition, they were instructed to

> decide which response is the answer choice that you think would make the best impression. Please answer the questions as if you were taking part in a college admission exam. We are not asking how you would really handle the situation, but rather how you could get the highest scores.

The total testing time was approximately 2 hours. A couple of weeks later, the participants were debriefed and received feedback about their results.

Predictor Measures

Cognitive ability test. To assess cognitive ability, the Advanced Progressive Matrices (APM; Set II; Raven, Raven, & Court, 1998), a widely used measure of higher order general mental ability in both research and applied settings, was administered. The APM consists of 36 items, with a time limit

of 40 minutes. The test booklet indicated that all test-retest reliabilities were above .90. The construct validity of the APM has been supported by its high correlation (.74) with the Wechsler Adult Intelligence Scale.

Personality measure. The Big Five personality dimensions (neuroticism, extroversion, openness to experience, agreeableness, and conscientiousness) were measured using the authorized Flemish translation (Hoekstra, Ormel, & De Fruyt, 1996) of the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992). The NEO-FFI consists of 60 items, and each personality dimension is measured by 12 Likert-type items on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). A factor analysis (with principal axis extraction and varimax rotation) performed on our data resulted in five factors (with eigenvalues ranging from 5.1 to 2.3) that explained 32% of the variance. Fifty-two of the 60 items had their highest loading on the factors they purported to measure. In addition, the internal consistencies of the personality scores ranged from .67 (agreeableness) to .86 (neuroticism; Table 1). These values are generally in line with those reported elsewhere (Costa & McCrae, 1992). This enabled us to compute a score per subject on each of the five factors. This score was the mean self-rating on the scales, which belonged to a factor.

SJT. We used an SJT consisting of 23 items that presented situations about various student-related issues (teamwork, studying for exams, organizing, accomplishing assignments, interpersonal skills, social responsibility, perseverance, integrity) and asked students how they would respond. The scenarios were each followed by four alternative response behaviors. This SJT was the same as the SJT developed by Bess and Mullins (2002). First, we translated the 24 items of Bess and Mullins (2002) into Dutch, with a back-translation to English. One item had to be eliminated, and some items were slightly adjusted. An example item is presented in the Appendix.

Next, a scoring key was developed using subject matter experts (SMEs). Ten 1st-year graduate students, with a mean age of 23 years who were considered experts because they successfully completed their undergraduate studies, independently completed the SJT and indicated what they thought the most and least effective options were for each item. To ensure that the items were clearly understandable, the SMEs also made editorial comments and conservative changes to the content. Afterward, they gathered to compare their most and least effective options. The consensus had to be 80% or higher. If this was not the case, discrepancies were resolved through discussion.

The scoring method was similar to that of Motowidlo et al. (1990). If the participants chose the responses identified by the SMEs as best or worst, they received a score of +1 or –1 for each item, respectively, and they received a

Table 1
Means, Standard Deviations, Intercorrelations, and Reliabilities of Predictor and Criterion Measures

Variable	*M*	*SD*	Age	1	2	3	4	5	6	7	8	9
Age	20.61	2.32	—									
Predictor measures												
1. Cognitive ability	27.40	4.14	–.15**	—								
2. Neuroticism	35.78	8.17	–.09	–.08	(.86)							
3. Extroversion	40.87	5.45	–.03	.10	–.19***	(.67)						
4. Openness	44.42	6.18	.02	.11	.00	.12**	(.74)					
5. Agreeableness	42.95	5.39	.03	.02	–.05	.36***	.03	(.67)				
6. Conscientiousness	41.95	6.69	.07	.18***	–.18***	.19***	.09	.31***	(.81)			
7. SJT_H (n = 138)	11.30	4.27	.10	.06	–.10	.20**	.28***	.26***	.48***	(.62)	—	
8. SJT_F (n = 153)	15.05	3.37	.08	–.05	.01	–.02	–.03	.04	.18**	—	(.48)	—
Criterion measure												
9. GPA (n = 290)	594	107.78	–.09	.26***	–.03	.03	.07	.08	.28***	.33***	.09	—

Note. For the situational judgment test (SJT) only, participants were randomly assigned to an honest (SJT_H) or a fake (SJT_F) condition. Cronbach's α values are in parentheses. Grade point average (GPA) was measured on a scale ranging from 0 to 1,000, with higher scores indicating better grades.
**p < .05. ***p < .01.

score of 0 if their responses were one of the other two options. The total test score was created by summing item scores across the 23 items. This method resulted in a final score ranging from –23 to +23.

The internal consistency reliabilities of the SJT scores (.62 in the honest condition and .48 in the fake condition) were rather low. However, internal consistency is not an appropriate reliability coefficient (as opposed to test-retest reliability) because SJTs typically measure heterogeneous content (Chan & Schmitt, 1997; Clause, Mullins, Nee, Pulakos, & Schmitt, 1998; Motowidlo & Tippins, 1993).

Criterion Measure

Given that an SJT of college students' success was used, the criterion measure of college students' success was operationalized as a student's GPA (Roth & Bobko, 2000). In our study, GPA was the average of students' 1st- and 2nd-year GPAs. We gathered these data with university authorization. The GPAs of 290 participants (99%) were available. GPA was measured on a scale ranging from 0 to 1,000, with higher scores indicating better grades. The correlation between participants' 1st- and 2nd-year GPAs was .68. This value is similar to the values found in a meta-analysis on the temporal stability of GPA (Vey et al., 2003).

Results

Effect on Mean Scores

Descriptive statistics and correlations of all study variables are presented in Table 1. Alpha values are shown on the diagonal in parentheses. An independent-samples *t* test revealed that students in the fake condition had significantly higher SJT scores ($M = 15.05$, $SD = 3.37$) than students in the honest condition ($M = 11.30$, $SD = 4.27$), $t(289) = -8.36$, $p < .001$.

To assess whether the faking effect was practically significant, we computed the effect size, which was about one standard deviation ($d = .89$), with women ($d = .94$) being better able to fake on the SJT than men ($d = .76$). In addition, we examined the percentage of fakers who were in the highest quartile. This parallels the assumption that the selection ratio of students who could gain access to higher education was 0.25. In this quartile (an SJT score of 16 or higher), 76% were fakers and 24% were honest respondents. In the lowest quartile (an SJT score of 11 or lower), 69% were honest respondents and 31% were fakers. This shows that faking on an SJT might exert substantial practical effects on who is selected.

Table 2
Hierarchical Regression Analysis of Grade Point Average on Cognitive Ability, the Big Five Personality Dimensions, and Situational Judgment

	Condition							
	Honest (n = 138)				Fake (n = 153)			
	β	t	p	R^2 Increment	β	t	p	R^2 Increment
Step 1								
Cognitive ability	.13	1.54	.125	.024*	.31	3.97	.000	.091***
Step 2								
Neuroticism	.14	1.68	.095		–.08	–.98	.331	
Extroversion	.11	1.23	.221		–.20	–2.40	.018	
Openness	–.01	–.07	.941		.02	.20	.842	
Agreeableness	–.06	–.63	.533		.07	.78	.439	
Conscientiousness	.14	1.40	.164	.087**	.22	2.63	.009	.089**
Step 3								
SJT	.27	2.81	.006	.051***	.05	.70	.483	.003

Note. Parameter estimates are for final step, not entry. SJT = situational judgment test. R^2 = .162 and adjusted R^2 = .116 in honest condition. R^2 = .183 and adjusted R^2 = .143 in fake condition.
*p < .10. **p < .05. ***p < .01.

Effect on Criterion-Related Validity

The last row of Table 1 presents the criterion-related validity coefficients. Cognitive ability (r = .26, p < .01) and conscientiousness (r = .28, p < .01) had significant correlations with the GPA criterion. Inspection of the relationship between the different SJT conditions and our criterion shows that the correlation between the SJT and GPA was significantly larger in the honest group (r = .33, p < .001) than in the fake group (r = .09, *ns*), z = 2.15, p < .05. These results indicate that faking had a negative impact on the criterion-related validity of the SJT.

We also used Thorndike's direct range restriction formula to correct the correlation of .09 in the fake condition because this lower correlation could have been due to the restriction of scores (see Table 1). This formula (with u = 1.27) gave a corrected correlation of .11 in the fake condition, which was lower than .33 in the honest condition. Thus, the lower validity in the fake condition was not due to range restriction.

Effect on Incremental Validity

We also examined whether faking affected the incremental validity of the SJT over and above cognitive ability and personality (see Clevenger et al., 2001). To this end, two hierarchical regression analyses (for the honest and

fake conditions) were conducted in which GPA was regressed on cognitive ability, the Big Five personality dimensions, and the SJT. The results are presented in Table 2. In the honest condition, cognitive ability and personality explained 2.4% and 8.7%, respectively, of the criterion variance, $F(1, 135) = 3.30$, $p < .10$, and $F(5, 130) = 2.54$, $p < .05$. The SJT explained an additional significant portion of the variance, namely, 5.1%, $F(1, 129) = 7.89$, $p < .01$. In the fake condition, the SJT did not add incrementally to the prediction of GPA over cognitive ability and personality, which explained 9.1% and 8.9%, respectively, of the criterion variance ($\Delta R^2 = .003$, *ns*). These results indicate that faking had a negative impact on the incremental validity of the SJT over cognitive ability and personality.

Additional Analyses

As noted above, one of the possible explanations as to why SJTs might be fakable is related to the constructs underlying SJTs. Therefore, an exploratory part of our study consisted of assessing the constructs related to the SJT of college students' success. To this end, we inspected the correlations between SJT overall scores and scores on the measures of cognitive ability and personality only in the honest condition ($n = 138$). The seventh row of Table 1 shows that the SJT had a nonsignificant correlation with cognitive ability ($r = .06$, *ns*) but had significant correlations with four of the Big Five personality dimensions: conscientiousness ($r = .48$, $p < .01$), openness ($r = .28$, $p < .01$), agreeableness ($r = .26$, $p < .01$), and extroversion ($r = .20$, $p < .05$). So, these results indicate that our SJT of college students' success was not heavily *g*-loaded. Conversely, the SJT was mainly related to the Big Five personality constructs. As already mentioned, these correlational patterns might serve as a possible explanation of why our SJT was fakable.

To examine even further whether the fakability of the SJT could be moderated by its *g*-loadedness, we examined this at the individual SJT item level. To this end, we correlated the 23 items of the SJT with cognitive ability. We also computed each item's fakability (i.e., effect sizes per item). Next, we correlated the fakability of each item with that item's correlation with cognitive ability. This final correlation was not significant ($r = .26$, *ns*, $n = 23$). However, it should be noted that there was little variation among the individual SJT items in terms of their *g* loading. In fact, none of the SJT items was correlated significantly with cognitive ability.

Discussion

Recently, there has been increasing interest in supplementing traditional student admission procedures with SJTs. Our study examined an important outstanding issue, namely, whether students can intentionally distort or fake

their responses on SJTs. This issue needs to be examined before SJTs can be widely adopted in undergraduate and graduate testing.

The first interesting finding of our study was that faking had a significant effect on SJT performance. This result indicates that if instructed to do so, students can fake an SJT of college students' success to make a favorable impression and to get higher test scores. Hence, fakers would be more able to gain access to higher education opportunities than honest respondents because we noticed that fakers rose to the top of the score distribution. The fact that our result is practically significant is also marked by the considerably large effect size ($d = .89$), comparable with the fakability of personality measures (see Viswesvaran & Ones, 1999).

A second contribution of this study is that the criterion-related validity of our SJT was .33 (in the honest condition). This value is comparable with the estimated population validity of .34 reported in the meta-analysis of McDaniel et al. (2001). The criterion-related validity of our SJT was also higher than the validity of general mental ability and conscientiousness, which are both well-established factors of academic achievement (Anderson & Keith, 1997; Boekaerts, 1996; Busato, Prins, Elshout, & Hamaker, 2000; De Raad & Schouwenburg, 1996; Wolfe & Johnson, 1995). Additionally, the student SJT explained 5.1% incremental variance over and above cognitive ability and personality (in the honest condition). All of this provides evidence for the generalizability of the validity of SJTs to student selection, and it indicates that SJTs can serve as a useful supplement to traditional student selection procedures (Lievens & Coetsier, 2001; Oswald et al., 2004). This is likely because the item content of SJTs can be tailored to mimic real-life situations related to college success.

The third and probably most important finding of our study is that when students fake, and they probably do this in a selection context, our SJT is no longer a valid predictor because the correlation coefficient dropped significantly to .09, which was not due to range restriction. Thus, on the basis of our study, faking negatively affected the criterion-related validity of the SJT. Furthermore, faking had a negative impact on the incremental validity of the SJT over and above cognitive ability and personality because the SJT no longer explained incremental variance. These disappointing results suggest that faking might be a possible threat to the use of SJTs in high-stakes testing programs. Clearly, future research is needed to confirm this result for other SJTs and in other populations. In a similar vein, future studies should examine whether SJTs are susceptible to coaching effects (Clause, Delbridge, Schmitt, Chan, & Jennings, 2001; Ryan, Ployhart, Greguras, & Schmit, 1998). As SJTs become more popular, test preparation firms will teach people how to respond to them most effectively. We still do not know whether coaching is a possible threat to the use of SJTs in a student admission context (Sackett, 2002).

In terms of future faking (and coaching) research related to SJTs, it is of key importance that we know why and under which conditions SJTs are fakable. This is because answers to this question might provide general guidelines to make SJTs more resistant to faking. According to the first possible explanation, SJTs are fakable depending on the instructions used. Specifically, Nguyen et al. (2002) found that knowledge-related instructions ("What should you do?") were more resistant to faking than behavioral tendency instructions ("What would you do?"). The second related explanation might be that faking depends on the constructs measured by the SJT. For example, Juraska and Drasgow (2001) suggested that the constructs assessed in an SJT could influence the level of fakability because some constructs are more fakable than others. The general idea is that an SJT is less fakable when it is *g*-loaded (Nguyen et al., 2002). Conversely, it is more fakable when it correlates more strongly with personality traits. The results of our study point in that direction. In fact, our SJT was fakable, and it had significant correlations with four of the Big Five personality dimensions. Moreover, our SJT correlated with cognitive ability at neither the overall score level nor the individual item level. Yet another explanation is related to the fact that SJTs are so-called low-fidelity simulations (Motowidlo et al., 1990). It is to be expected that these low-fidelity simulations might be more prone to faking than high-fidelity simulations, such as assessment centers, because the low-fidelity format of an SJT requires people only to choose one of the listed alternatives instead of acting out the response.

Future research is needed to test these alternative explanations. For example, one could compare an SJT that is significantly *g*-loaded with one that is not *g*-loaded, or one that measures mainly cognitive ability with one that measures mainly personality. Another possibility is to scrutinize the cognitive loading of SJT items. Next, the items can be grouped on the basis of their cognitive loading and then be compared on their fakability. Finally, it should be noted that a verbal protocol analysis of respondents' answering strategies and thought processes might shed an interesting light on why and when people fake.

Our study is not without limitations. First, our study adhered to the experimental faking research and therefore was a worst-case scenario, compared with an actual selection situation. This also means that our study examined the questions "Can students fake an SJT?" and "Can faking adversely influence the validity of an SJT?" These are different questions than "Do students in fact engage in faking SJTs" and "Does faking negatively affect the validity of an SJT?" Hence, future studies are needed to confirm our findings in a field setting in which students complete an SJT in an actual admission context. In addition, our findings are based on a sample of psychology students and therefore probably suffer from range restriction.

The second limitation is related to the criterion measure used. The criterion in the present study was GPA because past studies have shown that GPA

is a valid measure of academic success (Feldhusen & Jarwan, 1995). Furthermore, our results were in line with the GPA literature because GPA was correlated with two well-established factors of academic achievement, namely, general mental ability and conscientiousness (Anderson & Keith, 1997; Boekaerts, 1996; De Raad & Schouwenburg, 1996). Yet, it is clear that students' performance also encompasses aspects other than GPA (e.g., interpersonal skills, leadership, perseverance; Oswald et al., 2004). Therefore, future research is needed to use other outcome measures or performance dimensions besides GPA.

Third, it is important to note that SJTs are externally constructed measures. SJTs are developed for a specific criterion because their items are directly developed or sampled from the criterion behaviors that the SJTs are designed to predict (Chan & Schmitt, 2002). In a similar vein, the SJT scoring key is based on a specific criterion. Although this increases the contextualized nature and the realism of the SJT, it also leads to a drawback. Probably, the validity of our specific SJT of college students' success and its expert scoring key will not generalize across a wide variety of situations (e.g., different universities, different subject matter fields, universities in different countries, different types of curricula). This contrasts with internally constructed measures, such as cognitive ability measures, whose validity has been found to generalize across a wide variety of jobs and occupations (Schmidt, 2002).

Taken together, our study shows that students can fake an SJT of college students' success if instructed to do so. Faking had a negative impact on both the criterion-related validity and the incremental validity of the SJT over and above cognitive ability and personality. Additionally, we have outlined possible future research avenues that must be considered before SJTs can become a useful supplement to traditional student selection procedures.

Appendix
Example Item of a Situational Judgment Test of College Students' Success

You have so many assignments to complete and so much studying to accomplish, you feel you will never get caught up or accomplish anything. You are truly overwhelmed. What would you do?

a. Prioritize your activities, enumerate the steps to be accomplished for each activity, and systematically go through your work.[a]
b. Decide what you can accomplish reasonably and focus on getting that work done, and let the rest of the work unfinished.
c. Talk to your professors, explaining your situation, and ask for extensions on the due dates.
d. Take a break for a day and go out with your friends, then go back to working hard again.

a. Correct or best answer.

References

Abrahams, N. M., Neumann, I., & Githens, W. H. (1971). Faking vocational interests: Simulated versus real life motivation. *Personnel Psychology*, *24*, 5-12.

Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, *60*, 59-72.

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.

Anderson, E. S., & Keith, T. Z. (1997). A longitudinal test of a model of academic success for at-risk high school students. *Journal of Educational Research*, *90*, 259-268.

Atkinson, R. C. (2001, April). *Standardized tests and access to American universities*. Lecture delivered at the 83rd annual meeting of the American Council on Education, Washington, DC.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking on a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389-406.

Bess, T. L., & Mullins, M. E. (2002, April). *Exploring a dimensionality of situational judgment: Task and contextual knowledge*. Paper presented at the 17th annual conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Boekaerts, M. (1996). Personality and the psychology of learning. *European Journal of Personality*, *10*, 377-404.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, *29*, 1057-1068.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*, 143-159.

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*, 233-254.

Clause, C. S., Delbridge, K., Schmitt, N., Chan, D., & Jennings, D. (2001). Test preparation activities and employment test performance. *Human Performance*, *14*, 149-167.

Clause, C. S., Mullins, M. E., Nee, M. T., Pulakos, E., & Schmitt, N. (1998). Parallel test form development: A procedure for alternative predictors and an example. *Personnel Psychology*, *51*, 193-208.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, *86*, 410-417.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psycholocical Assessment Resources.

Cowles, M., Darling, M., & Skanes, A. (1992). Some characteristics of the simulated self. *Personality and Individual Differences*, *13*, 501-510.

Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid report. *Journal of Applied Psychology*, *79*, 643-658.

Dalen, L. H., Stanton, N. A., & Roberts, A. D. (2001). Faking personality questionnaires in personnel selection. *Journal of Management Development*, *20*, 729-742.

De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality*, *10*, 303-336.

Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996, August). *The validity of non-cognitive measures decays when applicants fake*. Paper presented at the annual conference of the Academy of Management, Cincinnati, OH.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology*, *15*,13-24.

Dwight, S. A. (1999). *An assessment of the different effects of warning applicants not to fake*. Unpublished doctoral dissertation, State University of New York at Albany.

Edens, P. S., & Arthur, W. (2000, April). *A meta-analysis investigating the susceptibility of self-report inventories to distortion*. Paper presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, *86*, 122-133.

Feldhusen, J., & Jarwan, F. (1995). Predictors of academic success at state-supported residential schools for mathematics and science: A validity study. *Educational and Psychological Measurement*, *55*, 505-512.

Furnham, A. (1986). Response bias, social desirability & dissimulation. *Personality and Individual Differences*, *7*, 385-340.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology*, *16*, 573-592.

Griffith, R. L., Frei, R. L., Snell, A. F., Hamill, L. S., & Wheeler, J. K. (1997, April). *Warning versus no-warnings: Differential effect of method bias*. Paper presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Haas, A., & McDaniel, M. A. (1999, April). *Faking strategies: Effects on a situational judgment test*. Paper presented at the 14th annual conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Hezlett, S. A., Kuncel, N. R., Vey, M. A., Ahart, A. M., Ones, D. S., Campbell, J. P., et al. (2001, April). *The predictive validity of the SAT: A meta-analysis*. Paper presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO persoonlijkheidsvragenlijsten NEO-PI-R en NEO-FFI* [Manual for NEO personality inventories NEO-PI-R and NEO-FFI]. Lisse, the Netherlands: Swets & Zeitlinger.

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

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.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*, 152-194.

Hough, L. M., & Schneider, R. J. (1996). The frontiers of I/O personality research. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 31-88). San Francisco: Jossey-Bass.

Juraska, S. E., & Drasgow, F. (2001, April). *Faking situational judgment: A test of the Conflict Resolution Skills Assessment*. Paper presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Kluger, A. N., & Collela, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology*, *46*, 763-780.

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology*, *76*, 889-896.

Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology*, *76*, 698-707.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 162-181.

Lievens, F., Buyse, T., & Sackett, P. R. (in press). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology.*

Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, *10*, 245-257.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, *32*, 247-256.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*, 882-888.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*, 730-740.

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

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

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*, 640-647.

Motowidlo, S. J., Hanson, M. A., & Crafts, J. L. (1997). Applied measurement methods in industrial psychology. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 241-260). Palo Alto, CA: Davies-Black.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, *66*, 337-344.

Nguyen, N. T., & McDaniel, M. A. (2001). *Constructs assessed in situational judgment tests: A meta-analysis*. Paper presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Nguyen, N. T., McDaniel, M. A., & Biderman, M. D. (2002). *Response instructions in situational judgment tests: Effects on faking and construct validity.* Paper presented at the 17th annual conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Nicholson, R. A., & Hogan, R. T. (1990). The construct validity of social desirability. *American Psychologist*, *45*, 290-292.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, *11*, 245-269.

Ones, D. S., Viswesvaran, C., & Korbin, W. (1995, May). *Meta-analyses of fakability estimates: Between-subjects versus within-subjects designs*. Paper presented at the 10th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

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.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187-207.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleiter & J. S. Wiggins (Eds.), *Personality assessment via questionnaire: Current issues in theory and measurement* (pp. 143-165). Berlin, Germany: Springer-Verlag.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.

Ployhart, R. E., & Ehrhart, M. G. (2001, April). *Effects of response instructions on the criterion-related validity, construct validity, and reliability of situational judgment tests*. Paper presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Ployhart, R. E., & Ryan, A. M. (1998). The relative importance of procedural and distributive justice in determining applicants' reactions. *Journal of Applied Psychology*, *83*, 3-16.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced progressive matrices manual* (1998 ed.). Oxford, UK: Oxford Psychologists Press.

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

Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, *85*, 399-406.

Ryan, A. M., Ployhart, R. E., Greguras, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, *51*, 599-621.

Sackett, P. R. (2002, January). *New measures: Issues of validity and diversity.* Presentation at the College Board Conference on New Tools for Admission to Higher Education, Washington, DC.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, *56*, 302-318.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187-211.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, *77*, 629-637.

Schmitt, N., Clause, C. S., & Pulakos, E. D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. *International Review of Industrial and Organizational Psychology*, *11*, 116-139.

Schmitt, N., Sackett, P. R., & Ellingson, J. E. (2002). No easy solution to subgroup differences. *American Psychologist*, *57*, 305-306.

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.

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

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist*, *50*, 912-927.

Vey, M. A., Ones, D. S., Hezlett, S. A., Kuncel, N. R., Vannelli, J. R., Briggs, K. H., et al. (2003, April). *Relationships among college grade indices: A meta-analysis examining temporal influences*. Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

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

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, *52*, 679-700.

Weiner, J. A., & Gibson, W. M. (2000, April). *Practical effects of faking on job applicant attitude test scores*. Paper presented at 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement*, *55*, 177-185.

Worthington, D. L., & Schlottmann, R. S. (1986). The predictive validity of subtle and obvious empirically derived psychology test items under faking conditions. *Journal of Personality Assessment*, *50*, 171-181.

Zickar, M. J. (1997, April). *Computer simulation of faking on a personality test*. Paper presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, *84*, 551-563.