

The Effect of Warning Against Faking on Noncognitive Test Outcomes: A field Study of Bus Operator Applicants

Joni Kuroyama
Lamorinda Consulting, LLC

Chris W. Wright
San Francisco State University

Todd M. Manson
Indiana University Southeast

Chris J. Sablinski
California State University, Sacramento

The effects of a warning statement on predictor and faking scale scores of a noncognitive measure were examined in a field sample of Bus Operator applicants in a municipal transit agency. The experimental group completed the measure after receiving a warning that contained both identification and consequence elements. The control group received no warning. Applicants who were warned scored significantly lower on the two predictor scales, and were more likely to be categorized as "Honest" by the faking scale. The magnitude of the effects, however, was not sufficient to impact decisions regarding applicant advancement in the selection process. Implications for the use of warning statements in selection contexts and for future research are discussed.

The use of noncognitive measures in personnel selection is supported by extensive research indicating that they are valid predictors of job performance (Bliesener, 1996; Hunter & Hunter, 1984; Hurtz & Donovan, 2000; Mumford & Stokes, 1992; Reilly & Chao, 1982; Salgado, 1999; Schmitt, Gooding, Noe, & Kirsch, 1984; Schmidt & Hunter, 1998). However, concerns have also been raised about the susceptibility of these measures to applicant faking (Anderson, Warner, & Spencer, 1984; Becker & Colquitt, 1992; Donovan, Dwight, & Hurtz, 2003; Kluger, Reilly, & Russell, 1991; McFarland & Ryan, 2000; Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007; Schrader & Osburn, 1977). This concern led researchers to explore numerous ways to reduce or mitigate intentional response distortion, including imbedding faking scales, measuring item response latencies, applying statistical techniques, utilizing forced-choice

response formats, and requiring applicants to provide written elaborations of their answers (Ellingson, Sacket, & Hough, 1999; Holden, 1998; Heggestad, Morrison, Reeve, & McCloy, 2006; Schmitt & Kuncze, 2002).

One technique that has shown promise for reducing faking in organizational settings is the presentation of a warning statement prior to test administration (Dilchert, Ones, Viswesvaran, & Deller, 2006; Donovan et al., 2003; Dwight & Donovan, 2003; Kluger & Colella, 1993; Nordlund & Snell, 2006; Ramakrishnan & Snell, 2006; Rothstein & Goffin, 2006). Warning statements are explicit admonishments that urge applicants to respond truthfully. They may be formulated in a variety of ways, but research has generally focused on three types of statements. The first, termed identification, states that faking will be identified through the use of a faking index or lie scale items, or that responses will be verified in some manner. The second type, a consequence statement, suggests that faking will result in adverse consequences such as invalidation of test scores or disciplinary action. The third variation of the warning statement includes both identification and consequence elements and informs applicants that faking can be detected and will result in negative outcomes (Dwight & Donovan, 2003). Warnings are an attractive strategy because they are easy to incorporate into test administration and are cost-effective. More importantly, there are data to suggest that the use of warning statements may be useful in reducing applicant faking.

To date, the majority of the published research has focused on identification warnings. For example, research on the self-appraisal of general abilities such as intelligence and scholastic aptitude has indicated that identification warnings lead to higher accuracy and less leniency in self-appraisal ratings (Farh & Werbel, 1986; Mabe & West, 1982). Doll (1971) examined the effects of two types of identification warnings on biodata response faking among Aviation Officer Candidates. In one warning condition, applicants were told that the biodata questionnaire contained a lie scale, while those in another condition were told that their responses might be verified in a follow-up interview. Results showed that response distortion for applicants in both warning conditions was significantly lower compared to applicants in a third condition who were told to respond in a way that would make them look "as good as possible." Similar results were reported by Schrader and Osburn (1977) with a sample of college students who were told to assume that they were applying for either a sales assistant or general managerial position. Instructions indicating that a biodata measure contained a lie scale reduced applicant faking significantly.

Fox and Dinur (1988) warned an experimental group of males applying for a military training program that their responses to a self-assessment measure of nine different characteristics would be cross-checked with ratings made by superior officers and peers. Resulting correlations between self and supervisor/peer ratings were higher in magnitude across all of the nine measured

dimensions than correlations from a control group of applicants who were given no warning. Finally, Kluger and Colella (1993) warned half of a sample of nurse's assistant applicants that the scoring system for a biodata test would detect dishonest responses. The rest of the applicants did not receive a warning. Results showed that the warning alleviated faking, but only for certain types of items. Transparent items, which were presumably more easily faked, had means that were about one third of a standard deviation higher among applicants who were not warned, compared to those who were. Item variances were also smaller in the no warning condition, suggesting that these applicants tended to provide uniformly inflated responses to the items. There was no difference in item means for nontransparent items in the warning and no warning conditions, however, and item variances were reduced in the warning condition.

Research on the impact of consequence warnings and combined identification and consequence warnings is more limited. Regarding consequence warnings, Trent, Atwater, and Abrams (as cited in Dwight & Donovan, 2003) reported that the use of a consequence warning on a biodata measure for military recruits resulted in significantly lower predictor scores. Similar results were found by Wheeler, Hamill, and Tippins (as cited in Dwight & Donovan, 2003) with a personality measure administered to job applicants.

McFarland (2003) utilized a combined identification and consequence warning statement with a sample of college students. Scores on the Extroversion, Conscientiousness, and Neuroticism scales of the NEO-FFI were significantly lower for participants in the warning condition, as were scores on a measure of social desirability. In a similar study, Vasilopoulos, Cucina, and McElreath (2005) examined the influence of a combined identification and consequence warning statement on personality test item response latencies. Response latencies refer to the time between item presentation and item response. The premise is that response latencies are greater among applicants who are warned against faking because the warning makes the response decision more cognitively complex. The respondent must consider both their true response and the desirable response to the item, compare and weigh the consequences of each choice, and then choose a response. The results confirmed their hypothesis, as the use of the warning statement was associated with slower response latencies across six personality measures in a sample of federal law enforcement applicants.

The most comprehensive study to date on the effectiveness of various warning statements was conducted by Dwight and Donovan (2003). College students completed a variety of noncognitive measures and three faking indices under four different conditions: unwarned (UW), warning of identification of faking (WI), warning about consequences of faking (WC), and warning of both identification and consequences of faking (WB). Compared to the unwarned participants, personality scale scores were lower in the WI, WC, and WB conditions. Although the difference was statistically significant only for the WB

participants, participants in the warning conditions also scored lower on two of the three faking measures. The authors also noted the practical implications of their findings. If their data had been used in an actual selection process, approximately 21% of unwarned participants would have been screened out for faking, compared to less than 10% of participants in each of the three warning conditions.

In an attempt to summarize the available data, Dwight and Donovan (2003) conducted a meta-analysis of warning studies involving noncognitive predictors. Their results showed that, overall, warnings resulted in lower predictor scores ($d = .23$), but the effect was moderated by the type of warning given. Identification warnings had almost no effect ($d = .01$), while consequence warnings ($d = .30$) and combined identification and consequence warnings ($d = .25$) produced small to moderate effects. However, the authors commented that the results should be interpreted cautiously because only 15 studies were included in the meta-analysis.

The Present Study

While it is apparent from the literature that warning statements reduce applicant faking, it seems that the magnitude of the effect may be stronger with certain types of warnings. Concerning the latter point, there is not enough accumulated research to draw firm conclusions or to recommend certain types of statements over others. More importantly, there is only one published noncognitive measure/warning study that uses job applicants (Kluger & Colella, 1993). Thus, the goal of the present study was to advance the warning literature by evaluating the impact of a warning statement on noncognitive test scores in an employment setting. The authors were granted access to an organization that was piloting a commercially developed noncognitive instrument that contained an imbedded faking scale. Due to anticipated sample size issues and organizationally imposed constraints, the study design was restricted to the use of one type of warning statement. We chose a combined identification and consequence statement for two reasons. First, the combined statement had demonstrated small to moderate effects in similar studies. Second, because actual job applicants were used, it seemed more ethically and practically appropriate to warn applicants of both the mechanism for identifying faking and the consequences of doing so. Only informing applicants that faking could be detected would leave them unaware of the actions that would be taken against those identified as faking. Similarly, informing applicants of only the consequences of faking might raise questions about how faking was actually detected. It might also lead to suspicion about whether there actually was a faking detection mechanism, or if the employer was simply using a warning statement to encourage honesty. We concluded that informing applicants of both the presence of a faking scale and the

possibility that its results would be used to make hiring decisions was most appropriate for an actual selection context because it provided them with the most information about the process and how the results of the test would be used. Based on the findings of the current warning literature, the following hypotheses were developed.

Hypothesis 1: Mean predictor scores on the noncognitive measure will be significantly lower for applicants who receive a warning against faking than those who receive no such warning.

Hypothesis 2: The percentage of applicants identified as honest will be significantly higher for applicants who receive a warning against faking than for those who receive no such warning.

We were also interested in the practical impact of the warning statement if the noncognitive measure were to be used to make final applicant screening decisions. If, as assumed, warning statements result in lower predictor scores and more applicants being identified as honest by the imbedded faking scale, then a corresponding difference in the two applicant groups should be evidenced in selection process advancement. Specifically, applicants who receive warnings should be less likely to earn overall test scores that, based on the test developer's norms, make them eligible for hire. Because faking scale results would most likely be used to eliminate applicants identified as faking at this point a real selection process, we will use only applicants identified as honest to test this hypothesis.

Hypothesis 3: Among all applicants identified as honest, the percentage of applicants recommended for advancement in the selection process will be significantly lower for applicants who receive a warning against faking than for those who receive no such warning.

Method

Participants

The sample consisted of 196 applicants for the position of Bus Operator with a municipal transit agency in the Western United States. 55.1% of the applicants were male and 44.4% were female. One applicant (.5%) did not provide gender information. The ethnic composition of the sample was 72.4% African-American, 11.2% Asian/Pacific Islander, 7.7% Hispanic, 5.6% Caucasian, and 2.6% "Other." One applicant (.5%) did not report his/her ethnicity.

Measures

Bus operator selection survey plus (BOSS Plus). The BOSS Plus was developed specifically for transit operator selection (EB Jacobs, 2006; Jacobs, Conte, Day, Silva, & Harris, 1996). It is a 75-item survey designed to measure variables that are predictive of bus operator performance. The BOSS Plus utilizes a variety of items that measure behaviors, attitudes, and behavioral intentions across 10 areas. The test manual indicates that items for 7 of the 10 areas (49 of the 75 items) were developed from a general biodata scale and include safety, attendance, customer service, temperament, timeliness, social involvement, and self-confidence. Items for the remaining 3 areas measure safe driving behavior, hazardous thought patterns, and time urgency. The developers report internal consistency (coefficient alpha) estimates ranging from .65 to .90 across the 10 item sets. A concurrent, criterion-related validity study with a sample of 864 bus operators from across the U.S. and Canada was utilized to validate the BOSS Plus against supervisor ratings and archival data related to workplace attendance, safety, and customer service.

BOSS Plus results are sent to the publisher for scoring, and the user is provided with predictor data for three separate scales: Attendance, Safety, and Customer Service. Attendance and Safety scores are provided in the form of raw scores. Customer Service data are presented categorically, with applicants scored as “High” or “Low”. Based on its validation data, the publisher recommends that the Attendance and Safety scores be used to screen applicants for the next phase in the selection process. To facilitate score interpretation, applicants are placed into one of five score groups based on their Attendance and Safety scores. The categories are “1 - Best Bet Candidates”, “2 - Potentially Strong Candidates: Attention to Safety Issues”, “3 - Potentially Strong Candidates: Attention to Attendance Issues”, “4 - Potentially Poor: Further Evaluation”, and “5 - Low Probability of Success.” In terms of score category interpretation and use, the publisher suggests that applicants in score groups 1 through 3 be considered for advancement in the selection process.

In addition to the three predictor scores, the BOSS Plus scoring profile provides an “Honesty” score. This score purports to identify individuals who “exhibit a response pattern that raises their Attendance and Safety scores regardless of their merit for high scores.” Honesty scores are reported in three categories: “Honest – appears to be honestly representing personal characteristics,” “Doubtful – appears to be slightly exaggerating positive personal characteristics,” and “Faking – appears to be highly exaggerating positive personal characteristics.”

Procedure

The BOSS Plus was administered to applicants for the position “Full-Time Bus Operator” from April 25, 2006 to September 12, 2006. The sample of applicants used for this study were all at the same stage in the organization’s selection process, having previously passed several hurdles including an application screen, a paper and pencil minimum skills test, a video-based test of driving skills, and a video-based situational judgment test of customer service skills. Before administration of the test, applicants were randomly assigned to one of two conditions. In the warned condition, applicants read the standard test instructions provided by the publisher plus the following statement, “Please be aware that responding in a manner which misrepresents your true self can and will be detected. This information will be used when determining your eligibility for employment.” The instructions and warning statement were also read aloud by the test administrator. As the applicants were about to begin the survey, the administrator made this additional statement, “Please be aware that there is a faking index encoded in the survey. In other words, the survey can determine that a person is faking their answers, therefore it is in your best interest to give honest responses.” In the not warned condition, applicants read the standard test instructions. As in the warned condition, the instructions were also read aloud by the test administrator. Test results for all applicants were sent to the publisher for scoring. Because this was a pilot test, BOSS Plus results were not used to make screening or hiring decisions.

Results

Of the 196 applicants, 102 were assigned to the warned condition and 94 were assigned to the not warned condition. Demographic breakdown of the applicants, by experimental condition, is provided in Table 1.

Hypothesis 1 was supported. Independent Samples T-Tests indicated that Attendance scores for applicants in the warned condition ($M = 40.39$, $SD = 8.84$) were significantly lower than those in the not warned condition ($M = 43.13$, $SD = 6.64$, $t(194) = 2.46$, $p < .05$, Cohen’s $d = .35$) and Safety scores for applicants in the warned condition ($M = 28.87$, $SD = 6.66$) were significantly lower than those in the not warned condition ($M = 30.55$, $SD = 4.68$, $t(194) = 2.06$, $p < .05$, Cohen’s $d = .30$).

Hypothesis 2, which stated that there would be a significant relationship between faking scale categorization and experimental condition, was supported. Applicants identified as “Honest” and “Faking” were compared. Those identified as “Doubtful” were eliminated from the analysis because the categorization did not definitively indicate whether they were or were not faking. A 2 x 2 Chi-

Square analysis showed a significant difference between the groups, $X^2(1, N = 147) = 4.80, p = .03, Cramer's V = .18$. Frequencies were lower than expected in the “Not Warned/Honest” and “Warned/Faking” conditions (*Adjusted Residuals* = -2.19) and greater than expected in the “Warned/Honest” and “Not Warned/Faking” conditions (*Adjusted Residuals* = 2.19). Cross-tabulation data are presented in Table 2.

Table 1
Demographic Statistics by Experimental Condition

Demographic variable	Warned condition		Not warned condition	
	<i>n</i>	%	<i>n</i>	%
Gender				
Male	56	55.4	52	55.3
Female	45	44.6	42	44.7
Total	101	100	94	100
Race				
Caucasian	6	5.9	5	5.3
African-American	72	71.3	70	74.5
Hispanic	8	7.9	7	7.5
Asian	12	11.9	10	10.6
Other	3	3.0	2	2.1
Total	101	100	94	100

Note. *n* = 102 in the warned condition. One applicant did not provide gender or race information.

Table 2
Observed Frequency of Applicants in Each Faking Scale Category by Experimental Condition

Faking scale category	Not warned	Warned	Total
Honest	30 (43.5%)	48 (61.5%)	78
Faking	39 (56.5%)	30 (38.5%)	69
Total	69	78	147

Note. Percentages represent column percentage values.

In order to test Hypothesis 3, applicants in both conditions whose faking scale scores identified them as “Faking” or “Doubtful” were removed from the analysis. From the remaining sample of 78, those whose Attendance and Safety subtest scores placed them in the top three scoring groups were re-coded as “1 – Screened In” and applicants in the bottom two scoring groups were re-coded as “0 – Screened Out.” This coding scheme conformed to the recommendations provided in the test manual for utilizing BOSS Plus scores to identify applicants for further screening (EB Jacobs, 2006). A 2 x 2 Chi-Square test of independence showed that there was no significant relationship between warning condition and score category, $X^2(1, N = 78) = .641, ns$. Hypothesis 3 was not supported. Cross-tabulation data are presented in Table 3.

Table 3
Observed Frequency of Applicant BOSS Plus Score Categorization by Experimental Condition

Score group	Not warned	Warned	Total
Screened in	14 (46.7%)	18 (37.5%)	32
Screened out	16 (53.3%)	30 (62.5%)	46
Total	30	48	78

Note. BOSS Plus score categorization was derived from a composite of Attendance and Safety scores, according to test publisher guidelines. Applicants who were identified as faking or doubtful by their faking scale score were removed prior to analysis. Percentages represent column percentage values.

Discussion

This study was designed to test the effects of a warning statement on predictor and faking scale scores of a noncognitive measure in a field setting. Results supported the hypothesis that warning applicants against faking would reduce predictor scores on the selection instrument. Warned applicants were also less likely to be identified as faking on the imbedded faking scale. The magnitude of the mean scale differences was moderate, with effect sizes comparable to previous meta-analytic data (Dwight & Donovan, 2003).

This study contributes to the literature in this area in several ways. The results provide additional evidence of the effects of a combined identification and

consequences warning statement on predictor and faking scale scores. Also, these effects were demonstrated in a sample of actual job applicants, unlike most published research in the area. To date, only one published noncognitive/warning study has utilized a sample of job applicants (Kluger & Colella, 1993) and most studies examining the effects of combined identification and consequences warning statements on personality tests have utilized student samples (e.g., Dwight & Donovan, 2003; McFarland, 2003).

The results of this study also have implications for selection practice because they add to the body of research that indicates that applicants are capable of faking their scores on noncognitive measures and that they will do so in an actual selection context (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Morgeson, et al., 2007). Furthermore, the results suggest that employers may be able to mitigate the impact of faking by providing warning statements to applicants. Providing oral or written warnings against faking on a noncognitive test is a simple and cost-effective strategy that organizations can use to influence applicants to provide more honest responses and to reduce their motivation to artificially inflate their test scores.

To illustrate the practical impact of the warning effect, if the test results from this study had been used as a screening instrument, 62% of applicants who received warnings would have remained eligible for consideration based on their faking scale scores, compared to 44% of those who were not warned. Among the remaining applicants, 38% of those who received warnings earned predictor scores that met the cutoffs for advancement in the selection process, compared to 47% of those who were not warned. Although this difference was not statistically significant, it was in the expected direction and demonstrates that the use of the warning statement would have influenced the composition of the applicant pool. It should also be noted that our study was conducted with only one group of applicants from a selection process that is administered on a continuous basis. If we assume that our sample was representative of typical applicants for the bus operator job, then we would predict that the impact of the warning effects on the applicant pool over several exam administrations would be substantial.

The ramifications of these effects are particularly important if an organization is concerned about applicant faking and utilizes a top-down selection strategy with a small selection ratio. Rosse, Stecher, Miller, and Levin (1998) found that applicants who had high faking scores were much more likely to be hired when the selection ratio was restrictive. When the selection ratio was based on the top 5% of applicants, for example, seven of the eight applicants who were hired had unusually high faking scores. Similarly, Weiner and Gibson (2000) found that when the selection ratio decreased, the percentage of high-faking applicants who were hired increased. Therefore, as applicants who faked their responses achieved scores that put them at the top of the distribution, it was more

likely that they would be hired if a top-down hiring strategy was used (Hough, 1998; Mueller-Hanson, Heggstad, & Thornton, 2003; Rosse et al., 1998).

It is important to note, however, that our results do not indicate whether warnings would have resulted in better hires. Because this was a pilot test, we were unable to obtain hiring data from the applicant sample and therefore could not determine the impact of the warning statement on future hiring decisions. We were also not able to collect job performance data in order to establish the criterion-related validity of the test or to determine if validity coefficients differed between experimental conditions.

Limitations and Future Directions

In addition to the lack of hiring and job performance data, there are two other limitations to our findings. First, the fact that data were collected in the course of an organizational selection process limited our ability to apply procedural controls during the study. Therefore, no manipulation checks were implemented to measure the salience of the written and verbal warnings. Second, only one type of warning, a mixed identification and consequence statement, was evaluated in this study. This limits the generalizability of our findings to other variations of warning statements. We were also unable to compare the effectiveness of different statements.

Future research should evaluate the effectiveness of different types of warning statements in actual selection contexts, as there is limited field research in this area. Where possible, data should also be gathered to determine how the use of warning statements impacts test validity. In one of the only field studies to date on the topic, Harold, McFarland, Dudley, and Odin (2004) found that the validity coefficient of the NEO-FFI for predicting customer service employee performance was significantly lower for a group of employees who were warned that the test contained a lie scale, compared to a group that received no such warning. The same pattern of results was found for scores on a social desirability measure, which were predictive of job performance in the no warning condition, but not in the warning condition. The authors explained their findings by stating that, “the ability to respond to situational cues and adapt one’s behavior accordingly may be a key component in effective customer service performance.” They also recommended against using warning statements with noncognitive assessment for customer service jobs. This recommendation is supported by research and theory in several work domains (Hogan, 1998; Hogan, Hogan, & Busch, 1984; Hogan, Hogan, & Roberts, 1996; Hogan & Shelton, 1998; Viswesvaran, Ones, & Hough, 2001). Customer service is an essential component of the job of bus operator, so it can be argued that the use of a warning statement to increase honesty or a faking scale to identify response distortion may actually

be counterproductive in this study. Without data to link faking scale scores to job performance, however, there is no evidence to support or refute these assertions.

Because the ability to distort responses to noncognitive items may be positively related to performance in some jobs, data should also be gathered to further explore the effects of faking on the utility of noncognitive instruments. Currently, findings are mixed. Some studies have found that faking does not affect criterion validity (Amelang, Schafer, & Yousfi, 2002; Barrick & Mount 1996; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ones & Viswesvaran, 1998; Schmitt & Oswald, 2006; Weiner & Gibson, 2000) while others have found that faking reduces criterion validity (Jackson, Wroblewski, & Ashton, 2000; Mueller-Hanson et al., 2003; Schmit & Ryan, 1992; Schmit, Ryan, Stierwalt, & Powell, 1995; White, Hunter, & Young, 2006).

The debate over the use of noncognitive predictors in personnel selection continues to be a salient topic for industrial/organizational psychologists (Hogan, Barrett, & Hogan, 2007; Morgeson, et al., 2007; Murphy, & Dzieweczynski, 2005). There are legitimate questions about their incremental validity and susceptibility to applicant faking. The results of this study contribute to the body of literature that suggests that practitioners should be concerned about response distortion and should consider it when determining whether to use a noncognitive selection instrument.

References

- Amelang, M., Schafer, A., & Yousfi, S. (2002). Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychologische Beitrage, 44*, 24-41.
- Anderson, C. D., Warner, J. L., & Spencer, C. E. (1984). Inflation bias in self-assessment examination: Implications for valid employee selection. *Journal of Applied Psychology, 69*, 574-580.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*(3), 261-272.
- Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45*, 389-406.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317-333.

- Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational & Organizational Psychology*, *69*, 107-120.
- Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, *48*(3), 209-225.
- Doll, R. E. (1971). Item susceptibility to attempted faking as related to item characteristic and adopted fake set. *Journal of Psychology*, *77*, 9-16.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, *16*(1), 81-106.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, *16*(1), 1-23.
- EB Jacobs, LLC (2006). BOSS & BOSS plus bus operator selection survey. State College, PA: EB Jacobs.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*(2), 155-166.
- Farh, J. L., & Werbel, J. D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology*, *71*(3), 527-529.
- Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology*, *41*, 581-592.
- Harold, C. M., McFarland, L. A., Dudley, N., & Odin, E. P. (2004). Personality and Faking Behavior: Does Warning Moderate Validity? Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.
- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, *91*(1), 9-24.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and personnel selection, *Journal of Applied Psychology*, *92*(5), 1270-1285.
- Hogan, J., Hogan, R. T., & Busch, C. (1984). How to measure service orientation. *Journal of Applied Psychology*, *69*, 167-173.
- Hogan, R. (1998). Reinventing personality. *Journal of Social & Clinical Psychology*, *17*, 1-10.
- Hogan, R., Hogan, J., & Roberts, B. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, *51*, 469-477.
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance*, *11*, 129-144.

- Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior and Personality, 13*(2), 387-398.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*(2/3), 209-244.
- Hough, L. M., 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*(5), 581-595.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-88.
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*(6), 869-879.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance, 13*, 371-388.
- Jacobs, R. R., Conte, J. M., Day, D. V., Silva, J. M., & Harris, R. (1996). Selecting bus drivers: Multiple predictors, multiple perspectives on validity, and multiple estimates of utility. *Human Performance, 9*(3), 199-217.
- Kluger, A. N., & Colella, 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. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76*, 889-896.
- Mabe, P. A. III, & West, S. G. (1982). Validity of self-evaluation ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280-296.
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment, 11*(4), 265-276.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*(5), 812-821.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in selection contexts. *Personnel Psychology, 60*, 683-729.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. III, (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*(2), 348-355.
- Mumford, M. D., & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background measures.

- In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology (Vol. 3)*, pp. 61–138. Palo Alto, CA: Consulting Psychologists Press.
- Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*, 343–358.
- Nordlund, M., & Snell, A. (2006). *Improving criterion-related validities with instructions to applicants: A good idea that actually worked*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, *11*(2/3), 245-269.
- Ramakrishnan, M., & Snell, A. F. (2006). *Do warnings lead to honest personality self-ratings? A comparison of different strategies*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures, *Personnel Psychology*, *55*, 1-63.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634-644.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*, 155-180.
- Salgado, J. F. (1999). Personnel selection methods. In C.L. Cooper & I.T. Robertson (eds). *International Review of Industrial & Organizational Psychology*. New York: Wiley.
- 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*, *12*(2), 262-274.
- Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, *77*(5), 629-637.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607-620.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 407-422.
- Schmitt, N., & Kunce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology*, *55*, 719-730.

- Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selected settings. *Journal of Applied Psychology, 91*(3), 613-21.
- Schrader, A.D., & Osburn, H.G. (1977). Biodata faking: Effects of induced subtlety and position specificity, *Personnel Psychology, 50*, 395-404.
- Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology, 90*(2), 306-322.
- Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection & Assessment, 9*, 277-289.
- Weiner, J. A., & Gibson, W. M. (2000). *Practical effects of faking on job applicant attitude test scores*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- White, L. A., Hunter A. E., & Young, M. C. (2006). *Social desirability effects on the predictive validity of personality constructs*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Author Information:

Chris Wright, Ph.D.
Associate Professor
Department of Psychology
San Francisco State University
(415) 338-1465
cwright@sfsu.edu