

DEPARTURES FROM LINEARITY IN THE RELATIONSHIP BETWEEN
APPLICANT PERSONALITY TEST SCORES AND PERFORMANCE
AS EVIDENCE OF RESPONSE DISTORTION

Doug Haaland

Neil D. Christiansen, Ph.D.

Central Michigan University

Paper presented at the 22nd Annual International Personnel Management Association
Assessment Council Conference, Chicago, IL.

Correspondence should be addressed to:

Neil D. Christiansen, Ph.D.
Department of Psychology
Central Michigan University
Mt. Pleasant, Michigan 48859

neil.d.christiansen@cmich.edu

There has been controversy with regard to whether faking represents a threat to the validity of personality measures since they were first suggested as an aid to organizational decision-making (Edwards, 1957; Meehl & Hathaway, 1946; Scott, 1915). Although a large body of research indicates that personality measures *can* be distorted when test-takers are motivated to do so (Borislow, 1958; Dunnette, McCartney, Carlson, & Kirchner, 1962; Orpen, 1971), recent reviewers have concluded this research is largely irrelevant to applied selection situations because it has relied primarily on laboratory studies and simulations (Hogan, Hogan & Roberts, 1996; Hough & Schneider, 1996). In essence, the criticism is that "real-world" data has not supported the generalizability of findings that suggest a decay in the validity of personality measures attributable to faking (Ones & Viswesvaran, 1997).

The problem in assessing the prevalence and effects of faking is that only in laboratory studies and simulations is it possible to be certain which individuals have distorted responses and which have not. In actual job applicant samples it is difficult to differentiate between those respondents who might have obtained high trait scores because of motivated distortion versus those that scored high simply because they possess favorable traits. As a compromise, social desirability or response validity scales have frequently been employed in field situations in an attempt to identify individuals who are likely to have distorted their responses to personality inventories (Meehl & Hathaway, 1946; Rosenzweig, 1934; Windor, O'Dell, & Karson, 1975). Field research has consistently failed to find any beneficial effect on validity when desirability scale scores are used to correct for distortion variance (Barrick & Mount, 1996; Christiansen, Goffin,

Rothstein, & Johnston, 1994; Ones, Viswesvaran, & Reiss, 1996), suggesting either that faking is not much of a problem or desirability scales are not very sensitive to applicant distortion.

Thus the controversy over the effects of applicant faking of personality measures continues. Due to different methodologies used in laboratory and field studies, direct tests of the generalizability of the results found in simulations have not been possible. The purpose of the present study was to examine real-world data for evidence consistent with what might be predicted based on the results of faking studies conducted in situations with less ecological validity. More specifically, simulations show that when some respondents distort their scores those individuals will rise to the top of the distribution and those scores will suffer a serious decay in validity. Should this also occur in actual job applicant samples, one would expect a departure from linearity in construct relationships across different ranges of personality test scores. On the other hand, if faking does not represent a problem, construct relationships should be similar across different ranges of scores.

Fakers Rise to the Top

At least two simulation studies have demonstrated that when responses to personality inventories that have been faked are added to a sample of more honest responses, those individuals with distorted scores rise to the top of the distribution. Zickar, Rosse, and Levin (1996) conducted a Monte Carlo simulation where item response theory was used to model faking. Their results indicated that even when a relatively small number of faked scores are added to a distribution, the percentage of fakers found at the top end accumulates rapidly. The faked scores in this demonstration, however, were not actual responses but had been generated for the purpose of the simulation. Douglas,

McDaniel, and Snell (1996) asked one student sample to respond to personality items in an honest fashion and another to respond as if they were trying to create as favorable an impression as possible. They then conducted a similar simulation to that done by Zickar et al. (1996) by adding the distorted scores to those of the honest sample, finding that more than half of the top ten scorers were from the faking condition when only 10% of the total sample had distorted. When the number of fakers in the total sample was increased to 25%, nine of the top ten were individuals from the faking condition.

Distortion Degrades Validity

Simulations have shown that faking has a harmful effect on the validity of personality trait scores. For example, Christiansen (1998) demonstrated that scores of individuals asked to complete a personality measure as if applying for a sales job contained approximately one-third the trait variance of scores from those who were instructed to respond honestly. Also consistent with response distortion resulting in a change in the construct validity of trait scores, correlations between personality measures and social desirability are typically larger under faking instructions (Diaz, Christiansen, & MacKay, 1997; Paulhus, Bruce, & Trapnell, 1995). The preceding observations being the case, it is not surprising that audits of construct validity using multitrait-multimethod approaches and factor analytic approaches have also found distortion to cause a decay (Christiansen, 1998; Douglas et al., 1996; Frei, Griffith, Snell, McDaniel, & Douglas, 1997).

Perhaps more of a concern, faking research using simulations has suggested that applicant distortion may be injurious to the prediction of performance. For example, Schmit and his colleagues found that the validity of a measure of conscientiousness used in the prediction of college GPA observed under normal instructions ($r=.25$) was destroyed

by instructions to respond as if applying for college admission ($r=-.03$; Schmit, Ryan, Stierwalt, & Powell, 1995). Douglas et al. (1996) obtained supervisory performance ratings for the employed students in their simulation and correlated both honest and distorted trait scores with performance. Reminiscent of the Schmit et al. (1995) results, the substantial correlations between the personality measures and performance in the honest condition ($r=.37$ for agreeableness and $r=.24$ for conscientiousness) were obliterated by instructions to create a favorable impression ($r=.00$ and $r=-.02$ for agreeableness and conscientiousness, respectively) .

Cognitive Ability and Distortion

Although generally suggesting of problems caused by faking, simulations have also suggested a possible mechanism whereby response distortion of personality measures may not represent as serious a threat in actual selection situations as might otherwise be thought. Specifically, some evidence exists that cognitive ability scores are more highly correlated with personality and personality-based integrity test scores under faking conditions than under honest conditions. For example, Mersman and Shultz (1997) found small correlations under normal instructions between a measure of general mental ability and trait measures of conscientiousness ($r=-.03$) and emotional stability ($r=.10$). However, when the trait scales were completed one week later under fake-good instructions the correlations with cognitive ability were larger (r 's of .24 and .27, respectively). Similarly, Alliger, Lilienfeld, and Mitchell (1996) found that the correlations between an intelligence test and scores on both an overt and covert integrity test were larger in a "fake-good" condition than in a "just take" condition. This suggests that

persons higher in cognitive ability may be more successful at elevating their personality test scores when they are motivated to do so.

Given the fact that cognitive ability is predictive of performance across jobs, it might be that those rising to the top of the distribution of personality scores possess more of another job-related attribute. If faking decreases trait variance but the measures become saturated with variance that is also predictive of performance, validity might not suffer. Taken together, this might lend some credence to arguments that faking happens but does not harm criterion-related validity to a large degree.

Hypotheses

Based on the propositions that applicants distorting responses to personality measures in a favorable direction are found disproportionately in the upper ranges of the score distribution and distorted scores are worse predictors of job performance, it was predicted that:

Hypothesis 1: The correlation between personality and performance for applicants in the upper half of the distribution of personality scores will be lower than that for applicants in the lower half of the sample. Further, the correlation between personality and performance will be lowest for applicants at the top end of the distribution.

Should these individuals have obtained higher personality test scores by systematically endorsing more socially desirable responses, one would expect that their scores would demonstrate more dependency with social desirability scores due to common situational demands. This is similar to the observation that a very strong response set results in an heightened level of consistency across self-report measures. Thus:

Hypothesis 2: The relationship between personality and social desirability for applicants in the upper half of the distribution of personality scores will be stronger than that for applicants in the lower half of the sample. Further, the relationship

between personality and social desirability will be strongest for applicants at the top end of the distribution.

Finally, if persons with higher cognitive ability perform better and those individuals are also those who inflate their personality scores the most, it would follow that:

Hypothesis 3: The correlation between personality and cognitive ability for applicants in the upper half of the distribution of personality scores will be higher than that for applicants in the lower half of the sample. Further, the correlation between personality and cognitive ability will be highest for applicants at the top end of the distribution.

METHOD

Participants and Procedure

The study participants were 442 recruits who attended a Midwestern state police academy between 1993 and 1996. Three-hundred and seventy-five were males (85%), the mean age was 26 years ($SD=3.38$), and most reported themselves as Caucasian (83%) with some educational experience beyond high school (94%).

All participants had successfully completed the state police employment process, having passed the Civil Service examination, a physical examination, a background check, and a personality measure intended to identify proclivity toward counterproductive behavior. Several administrations of each test were given and all participants had completed the selection hurdles no more than one year prior to the first day of employment at the academy.

Measures

Personality. The Inwald Personality Inventory (IPI; Inwald, 1992) was administered to all recruits prior to being offered a space at the academy. The IPI is a 310-item

instrument that was developed specifically to assist public safety and law enforcement agencies in selecting new officers, and is comprised of 25 trait scales. Preliminary analyses were conducted to identify those subscales that were predictive of performance. Four subscales (Absence Abuse, Depression, Family Conflicts, and Illness Concerns) had statistically significant ($p < .05$) zero-order correlations with performance. Raw scores were recoded to the criterion such that higher scores on each scale would yield higher predicted performance. Each was transformed to T-scores with the four scales being averaged to form the personality composite used as the measure of personality in the study. The total sample was divided into three subgroups: (a) those in the lower half of the distribution, (b) those in the upper half, and (c) top scorers (approximately 15%; $n=53$).

Performance. Performance was based on eight core dimensions important to performance in the academy: (a) planning and organizing, (b) communication skills, (c) initiative, (d) equipment use/maintenance, (e) interpersonal skills, (f) academic/classroom performance, (g) appearance, and (h) professional conduct. Performance was measured on dichotomous scales with participants being considered either *needing improvement* or *meeting expectations* along each dimension, and was assessed twice during the 16-week academy period (Weeks 8 and 16) by senior officers trained in the administration of the performance appraisal system. Dimension scores from the two performance assessments were summed to form the final measure of performance.

Social Desirability. The response validity scale from the IPI (i.e., the Guardedness scale) was used as the measure of social desirability. The IPI Guardedness scale is designed to "identify persons who have minimized shortcomings, denied faults, and

answered items in a 'socially desirable' direction (Inwald, Knatz, & Shusman, 1983). Coefficient alpha for the Guardedness scale has been previously reported as .82 in a normative sample of 1,050 police officer candidates (Inwald, 1992).

Cognitive Ability. The written portion of the state Civil Service examination was used as the measure of cognitive ability. The written Civil Service exam is a knowledge-based multiple choice, true-false formatted exam that contains three parts: (a) a reading comprehension section, in which examinees were asked to read a paragraph and then answer questions about it; (b) an accuracy of observation section, where examinees were presented with a scene and then asked questions about what they had observed; and (c) a writing test. Coefficient alpha for the written Civil Service examination has been computed by the state Civil Service commission to be .86 (Palmatier, 1996).

RESULTS

Descriptive statistics are presented separately in Table 1 for the total sample, those in the lower half and upper half of the personality test score distribution, and the top scorers. Persons in the upper half were rated significantly higher on the performance measure than were those in the lower half, $F(1,440)=7.48, p<.01$. Similarly, those with personality composite scores in the upper half of the distribution scored over half a standard deviation higher on the social desirability measure than did those whose scores were in the lower half, $F(1,440)=42.97, p<.01$. Although the upper half subgroup displayed a slightly higher mean cognitive ability score than did the lower half subgroup (50.56 versus 49.48), the difference was not significant, $F(1,440)=1.27, ns$.

When comparing the top scorers with the lower half, only the mean difference in social desirability was significant, $F(1,279)=26.18, p<.01$. The top scorers scored nearly

three-quarters of a standard deviation higher on social desirability than did those in the lower half of personality composite scores. The mean differences in performance and cognitive ability between these two subgroups were not significant [$F(1,279)=2.12, ns$ and $F(1,279)=2.01, ns$, respectively], although the means of the top scorers were slightly higher on both of the two measures.

The relationships between personality and performance, social desirability, and cognitive ability are presented for each subgroup in Table 2. Due to the fact that differences in observed correlations may simply be the result of differences in variances rather than substantive differences in construct relationships, the unstandardized regression coefficients (b) are also provided in Table 2. Conceptualized simply as the correlation between two variables multiplied by the ratio of their standard deviations (Cohen & Cohen, 1983), the unstandardized regression coefficient is more robust to the effects of range restriction and, therefore, represents the best estimate of the true linear relationship between two constructs. In addition, the correlations corrected for range restriction in the personality composite (r) using the total sample to estimate the unrestricted variance were also computed and are provided in Table 2.

As shown in Table 2, the positive relationship between personality and performance found in the total sample was mirrored in the lower half, but negligible in the upper half and substantially *negative* for the top scorers. Specifically, in the total sample, increases in personality scores were associated with higher performance ($b=.18, r=.18$). Similarly, within the lower half subgroup, persons obtaining higher scores on the personality composite performed better ($b=.21, r=.20$). However, within the upper half subgroup, increases in personality scores did not have a beneficial effect on performance ($b=.01,$

$r=.02$). Further, among the top scorers, individuals obtaining higher personality composite scores performed *worse* ($b=-1.34$, $r=-.80$). This clearly supports Hypothesis 1.

The positive relationship between personality and social desirability in the total sample ($b=.31$, $r=.31$) was less pronounced in the lower half ($b=.18$, $r=.19$), but stronger in the upper half ($b=.39$, $r=.37$) and much stronger among the top scorers ($b=1.14$, $r=.77$). Although the relationship between personality and social desirability was positive for all subgroups, the results indicated that the relationship was stronger for individuals in the upper half of the personality score distribution than for individuals in the lower half. Further, the relationship between personality and social desirability was strongest for top scorers. Supportive of Hypothesis 2, this suggests that those with higher scores on the personality measure were more concerned with presenting themselves favorably than were those with lower scores.

The modest positive relationship between personality and cognitive ability in the total sample ($b=.10$, $r=.10$) was reflected in both the lower half ($b=.13$, $r=.12$) and the upper half ($b=.11$, $r=.10$). Among the top scorers, however, the relationship was substantially *negative* ($b=-.73$, $r=-.66$). In other words, in the total sample, persons scoring higher on the personality composite scored slightly higher on the cognitive ability measure. Similarly, both for individuals in the lower half and the upper half of the personality composite distribution, increases in personality scores were associated with slightly greater cognitive ability. Among top scorers, however, persons scoring higher on the personality composite demonstrated *lower* cognitive ability. Hence, Hypothesis 3 was not supported.

To summarize, in the lower half of the sample, a one unit increase in personality score resulted in a .21 unit increase in performance and a .18 unit increase in social desirability. In contrast, in the upper half of the sample, a one unit increase in personality resulted in only a .01 unit increase in performance but a .39 unit increase in social desirability. Further, among the top scorers, a one unit increase in personality resulted in a 1.34 unit *decrease* in performance but a 1.14 unit *increase* in social desirability. Clearly, the differences in the construct relationships between top scorers and those in the lower half of the sample were quite large and suggest a substantial decay in validity of the personality test scores. For example, the per-unit increase in social desirability associated with a one unit increase in personality was *over six times as great* among applicants at the top end of the personality distribution than among applicants in the lower half of the distribution.

DISCUSSION

The results of this study using actual job applicants are consistent with those of simulations finding a decay in the validity of personality measures in the upper ranges of applicant trait scores. In this study, however, no effort was made to identify those individuals who had attempted to fake nor was there any attempt to manipulate test-takers' motivation to score well on the personality test. Instead, hypotheses were generated based on findings from simulation faking studies and applied to a "real-world" job applicant sample. The results were compatible with the notion that faking degrades validity.

Specifically, the positive relationship between the personality composite and performance found in the total sample was *near zero* in the upper half of the personality

score distribution, and was actually *negative* among the top 53 scorers. In other words, within this subgroup of the 53 top scorers, higher scores on the personality composite were associated with *lower* performance. On the other hand, the relationship between personality and social desirability was stronger in the upper half and much stronger among the top scorers, suggesting a possible mechanism for why the scores in the upper ranges of the distribution were worse predictors of performance: applicants with scores in the upper ranges tended to inflate their scores simply by selecting the most desirable items.

Results suggest that scores in the upper ranges of the personality score distribution are saturated with social desirability and that increases no longer relate to superior performance. Further, the results with regard to top scorers suggests that personality scores at the extreme high end of the distribution may actually relate to worse performance. There are two possible reasons for this. First, if true scores are lower than distorted observed scores and this was more prevalent at the very top of the distribution of high scorers, these individuals might be expected to perform worse than those within this subgroup with less extreme (and less distorted) scores. Second, individuals who are willing to engage in intentional distortion during the employment process may make worse police officers even as early on as the academy. Individual differences in the tendency to misrepresent oneself during the application process may be indicative of future behavior targeted more toward impression management than actual job success.

Caution is urged in the interpretation of these data to infer that personality measures are not useful as predictors of job performance in applicant samples. Numerous validity coefficients from job applicant samples have contributed to meta-analytic research displaying useful levels of validity *in spite of faking* (Barrick & Mount, 1991; Hough,

1992; Tett, Jackson, & Rothstein; 1991). It is also worth noting that the usefulness of a predictor for selection is not determined by the correlation within the top-scoring group, but rather on the validity in the total sample. Indeed, correcting for range restriction using IPI normative data to estimate unrestricted variance (Inwald, 1992) and for criterion unreliability (based on .80; Pearlman, Schmidt, & Hunter, 1980) would yield an operational validity of $r=.32$. Further, consider the fact that in the present study persons in the upper half of the personality score distribution significantly outperformed those in the lower half despite the near-zero relationship between personality and performance in this subgroup. In fact, this result is virtually guaranteed by the positive correlation in the total sample.

Several implications regarding organizational personality assessment are evident from this study. First, personality testing may be a more useful predictor at higher selection ratios. Cut scores at more favorable selection ratios may be problematic because it is within this range that distortion has the most deleterious effect. On the other hand, it seems unequivocal that those at the bottom of the score distribution have deficits in job-related attributes, and were either unable or unmotivated to misrepresent themselves. Consequently, personality measures may be better suited for selection out. Second, concurrent validation methods using incumbent samples may be ill-suited for determining the actual operational validity of personality inventories used to select applicants. Finally, the method used herein of examining subgroup differences in applicant samples may be useful for gauging the amount of validity decay attributable to response distortion.

TABLE 1
Descriptive Statistics of Study Variables by Subgroup

Subgroup	Personality		Performance		Social Desirability		Cognitive Ability	
	M	SD	M	SD	M	SD	M	SD
Total Sample (<i>n</i> =442)	50.00	10.00	50.00	10.00	50.00	10.00	50.00	10.00
Lower Half (<i>n</i> =228)	42.78	8.81	48.75	10.63	47.11	9.27	49.48	10.13
Upper Half (<i>n</i> =214)	57.70	3.00	51.33	9.11	53.08	9.85	50.56	9.85
Top Scorers (<i>n</i> =53)	61.91	1.60	51.10	10.28	54.39	9.56	51.62	8.37

Note. Scores on all variables were standardized such that in the total sample the mean (M) was 50 and the standard deviation (SD) was 10.

TABLE 2
Relationships between Personality and Performance, Social Desirability, and Cognitive Ability by Subgroup

Subgroup	Performance			Social Desirability			Cognitive Ability		
	<i>b</i>	<i>r</i>	Γ	<i>b</i>	<i>r</i>	Γ	<i>b</i>	<i>r</i>	Γ
Total Sample (<i>n</i> =442)	.18	.18	.18	.31	.31	.31	.10	.10	.10
Lower Half (<i>n</i> =228)	.21	.18	.20	.18	.17	.19	.13	.11	.12
Upper Half (<i>n</i> =214)	.01	.01	.02	.39	.12	.37	.11	.03	.10
Top Scorers (<i>n</i> =53)	-1.34	-.21	-.80	1.14	.19	.77	-.73	-.14	-.66

Note. *b*=unstandardized regression coefficient; *r*=observed correlation, Γ =correlation corrected for range restriction in the personality composite using the total sample to estimate the unrestricted variance.

REFERENCES

- Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7*, 32-39.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261-272.
- Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology, 82*, 62-78.
- Borislów, B. (1958). The Edwards Personal Preference Schedule and fakability. *Journal of Applied Psychology, 42*, 22-27.
- Christiansen, N. D. (1998). *Sensitive or senseless? On the usefulness of social desirability measures for making inferences concerning applicant response distortion*. Paper presented at the Thirteenth Annual Meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847-860.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

- Costa, P. T., Jr., & McCrae, R. R. (1992). *The Revised NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Diaz, J., Christiansen, N. D., & MacKay, D. (1997). *The usefulness of validity scales for identifying response distortion in personality assessment*. Paper presented at the Annual Meeting of the Southeastern Psychological Association, Atlanta, GA.
- Douglas, E. F., McDaniel, M. A., & Snell, A. F. (1996). *The validity of non-cognitive measures decays when applicants fake*. Paper published in the Academy of Management proceedings (August, 1996).
- 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.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Frei, R. L., Griffith, R. L., Snell, A. F., McDaniel, M. A., & Douglas, E. F. (1997). *Faking of non-cognitive measures: Factor invariance using multiple groups LISREL*. Symposium conducted at the Twelfth Annual Meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions. *American Psychologist, 51*, 469-477.
- Hough, L. M. (1992). The "big five" personality variables--construct confusion: Description versus prediction. *Human Performance, 5*, 139-155.

- Hough, L. M., & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In K. Murphy (Ed.), *Individual Differences in Organizations*. San Francisco, CA: Jossey-Bass.
- Inwald, R. E. (1992). *Inwald Personality Inventory technical manual (Revised Version 2A)*. New York: Hilson Research.
- Inwald, R., Knatz, H., & Shusman, E. (1983). *Inwald Personality Inventory manual*. New York: Hilson Research.
- Meehl, P. E., & Hathaway, S. R. (1946). The K Factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, *30*, 526-564.
- Mersman, J. L., & Shultz, K. S. (1997). *Individual differences in the ability to fake on personality measures*. Paper presented at the Twelfth Annual Meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.
- Ones, D. S., & Viswesvaran, C. (1997). *The influence of social desirability on integrity test validities: A meta-analysis*. Symposium conducted at the Twelfth Annual Meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.
- 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.
- Orpen, C. (1971). The fakability of the Edwards Personnel Preference Inventory in personnel selection. *Personnel Psychology*, *24*, 1-4.

- Palmatier, J. J. (1996). *The big five-factors and hostility in the MMPI and IPI: Predictors of Michigan state trooper job performance*. Unpublished doctoral dissertation, Michigan State University.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21*, 100-108.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65*, 373-407.
- Rosenzweig, S. (1934). A suggestion for making verbal personality tests more valid. *Psychological Review, 41*, 400-401.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607-620.
- Scott, W. D. (1915). The scientific selection of salesmen. *Advertising and Selling, 24*, 94-96.
- Tett, R. P., Jackson, D. N., & Rothstein, M. G. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.
- Windor, P., O'Dell, J. W., & Karson, S. (1975). New motivational distortion scales for the 16PF. *Journal of Personality Assessment, 39*, 532-537.

Zickar, M., Rosse, J., & Levin, R. (1996). *Modeling the effects of faking on personality instruments*. Paper presented at the Eleventh Annual Meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.