

Paper published in the *Academy of Management Proceedings* (August, 1996).

Winner of the Best Paper Award for the Human Resource Division.

## **THE VALIDITY OF NON-COGNITIVE MEASURES DECAYS WHEN APPLICANTS FAKE**

Elizabeth F. Douglas, Department of Psychology, University of Akron, Akron, OH 44325-4301  
Michael A. McDaniel, Department of Psychology, University of Akron  
Andrea F. Snell, Department of Psychology, University of Akron

### **ABSTRACT**

The effects of situationally-induced response bias on the reliability and validity of non-cognitive measures were investigated. Biodata and personality scales were administered under faking and honest conditions. Under conditions of faking, alpha reliabilities of scales increase, yet construct and criterion-related validity decreases.

### **INTRODUCTION**

This paper presents evidence for the proposition that the construct and criterion-related validity of non-cognitive tests is substantially harmed when respondents fake good. We argue that the renewed research interest in the validity of non-cognitive measures (Barrick & Mount, 1991; Ones, Viswesvaran, & Schmidt, 1993; Rothstein, Schmidt, Erwin, & Owens, 1990; Tett, Jackson, & Rothstein, 1991) gives too little attention to the potential problems of applicant faking on the validity of the measures. We review the response bias literature and then provide evidence that the validity of non-cognitive measures is degraded when applicants fake good.

Response bias is the most frequent criticism of personality inventories used in personnel selection (Ones, Viswesvaran, & Reiss, 1995). Response distortion has been described throughout the literature by terms such as social desirability, impression management, faking, intentional distortion, and self enhancement (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Lautenschlager, 1994; Ones, Viswesvaran & Korbins, 1995). Research has indicated that non-cognitive measures such as personality measures and biodata inventories can be subject to response bias due to faking (Hough et al., 1990; Kluger, Reilly, & Russell, 1991; Ones, Viswesvaran, & Korbins, 1995). For example, Hough et al., (1990) found that respondents were able to successfully distort their self-descriptions when instructed to do so. Kluger et al., (1991) also found that when instructed to fake, subjects were able to improve their scores on a biodata measure by nearly one standard deviation. Ones, Viswesvaran, and Korbins, (1995) used meta-analysis to determine whether or not subjects are able to intentionally fake responses to personality inventories. Their results showed that subjects, on average, are able to improve their scores through faking by nearly half a standard deviation.

Most research on faking has focused on measuring the degree of socially desirable responding and using social desirability scores to correct for faking. Christiansen, Goffin, Johnston, and Rothstein (1994) examined the effects of correcting for faking in the Sixteen Personality Factor

Questionnaire. The results of their study showed that correcting for faking is unlikely to affect the criterion-related validity of the measure. This conclusion is well-supported by Ones, Viswesvaran, and Reiss, (1995) who analyzed the cumulative literature on response bias to assess the impact of faking on the criterion-related validity of the Big Five personality factors. Their analyses presented evidence that correcting for faking with social desirability scales does not improve the criterion-related validity of personality scales. In addition, they showed that social desirability is related to stable personality constructs, specifically emotional stability and conscientiousness. They demonstrated that removing the effects of social desirability from personality scale scores would remove relevant variance from the measures. Based on this evidence, Ones, Viswesvaran, and Reiss (1995) concluded that socially-desirable responding is not a problem concerning the use of personality measures in personnel screening.

Although response bias related to faking is correlated with stable personality traits, under certain circumstances, it can be situationally induced. For example, it is well documented that people can fake when instructed to do so by an experimenter (Hough et al., 1990; Kluger et al., 1991; Ones, Viswesvaran, and Korbins, 1995). Thus, it is reasonable that people may also fake effectively when they are motivated to fake through self interest. For example, individuals may engage in faking behaviors, such as distorting resumes and answers in interviews and employment tests, when it is seen as necessary to get a valued outcome, such as employment. This was demonstrated by Stokes, Hogan and Snell (1993) where applicants were more likely to engage in impression management than incumbents due to motivation to increase their probabilities of obtaining employment. It is also supported by anecdotal evidence provided by staffing professionals who review applications and conduct interviews.

We agree that the evidence concerning faking has illustrated that social desirability is a stable personality trait related to individual differences in personality (Ones, Viswesvaran & Reiss, 1995). We also concur that correcting for the effects of faking does not improve the criterion-related validity of non-cognitive measures. However, we disagree with the assertion, often based on this evidence, that faking is not a problem concerning non-cognitive measures. We argue that faking, particularly situationally-induced faking, is a common problem in many testing situations. In this paper, we assess the extent to which faking serves to damage the construct and criterion-related validity of the instrument.

## **METHOD**

### **Sample**

The sample for this study consisted of 600 college students attending a Midwest university. Subjects were randomly assigned to one of the two conditions: 293 were placed into the honest condition and 307 were placed into a faking condition.

### **Measures**

A 64 item personality instrument consisting of single statement items requiring a true or false response was used. The instrument yields scale scores for conscientiousness and agreeableness. A

60 item, multiple choice, biographical data questionnaire was also administered. It also yields scale scores for conscientiousness and agreeableness. The four scales (two biodata and two personality scales) were developed using a construct-oriented approach (Hough & Paullin, 1994). Items in each of the four scales were scored rationally. Item analysis, conducted on the honest subjects, was used to refine the scales. Items with item-total correlations of less than .10 were discarded. A supervisor performance appraisal form, modeled after the form used by the U.S. Department of Labor for the validity studies of the General Aptitude Test Battery (GATB), was submitted to the subject's employer, with the subject's informed consent, to collect the criterion data for this study. Supervisors responded to two demographic questions and nine performance questions. A sample performance appraisal is available on request from the authors.

## **Procedure**

Subjects were randomly assigned to one of the two conditions. In the honest condition, subjects were asked to complete a biodata questionnaire and a personality measure. This group received the instructions to complete the questionnaires as honestly as possible. The second group, the faking condition, received both questionnaires but the instructions they received differed. The second group was told to answer the questions in such a way as to make as favorable an impression as possible. The instructions were:

*When completing this survey when applying for a job, many applicants stretch the truth in an attempt to make a favorable impression. When answering the questions, please try to make as favorable an impression as possible. Thus, we are not interested in your true opinion, rather we would like you to identify the most socially desirable response for each question.*

An example was given in the instructions in each condition to facilitate the understanding of the instructions.

Job performance ratings were solicited from the employers of the subset of subjects who worked and who provided permission for their supervisor to be contacted. Employers were sent a letter explaining the purpose of the study. A job performance appraisal form and a copy of the subject's signed informed consent sheet were also included. A response rate of 68% was obtained yielding 208 observations with criterion data.

Alpha reliability analyses were conducted for all four scales. Multi-trait and multi-method validity analyses and factor analyses were conducted to examine the construct validity of the scales. Criterion-related validities were also calculated. These sets of analyses were conducted for all subjects and separately for each of the two experimental conditions. Factor analyses were also conducted separately for each experimental condition. Criterion-related validities for the scales derived from both factor analyses were calculated for all subjects and separately by condition.

## RESULTS

### Reliability and Mean Difference Analyses

Table 1 presents the scale reliabilities (alpha) and the scale means and standard deviations for the combined conditions (statistics are based on all observations in both the honest and faking conditions) and separately by condition (honest or faking). The scores are standardized to make the scale scores in the honest condition have a mean of 50 and standard deviation of 10. An examination of the scale means show that subjects scored substantially higher in the faking condition than in the honest condition. The degree of faking ranged from .7 standard deviations of score improvement for the personality agreeableness scale to 1.3 standard deviations of score improvement for the biodata agreeableness scale.

-----  
Table 1 about here  
-----

The alpha reliabilities by condition, reported in Table 1, prove interesting. For 3 of the 4 scales, alpha reliabilities are higher in the faking condition than in the honest condition. We suggest that those who respond honestly report true inconsistencies in their own behavior. For example, with respect to conscientiousness-related behaviors, one might report that she arrives at work on time, but that she also does personal errands on work time. These true inconsistencies tend to reduce the homogeneity of the scale for the honest respondents. In contrast, those in the faking condition portray themselves as consistently performing positive behaviors and the alpha reliabilities reflect this consistency of response.

### Construct Validity Analyses

Table 1 presents 3 correlation matrices of the four rationally-keyed scales. The first matrix, titled *Combined Conditions*, is the matrix produced when the observations in the honest condition are pooled with the observations in the faking condition. The scales exhibit multi-trait and multi-method validity. That is, scales measuring the same constructs (e.g. conscientiousness) but with different methods (biodata vs. personality) correlate more highly with each other than do scales measuring different constructs but sharing the same measurement method. The multi-trait and multi-method validity is substantially better when the correlation matrix is calculated solely on the respondents in the honest condition. In contrast, the multi-trait and multi-method validity shows a decay in divergent coefficients when the correlation matrix is calculated solely on the subjects in the faking condition.

Exploratory factor analyses were also conducted separately by condition. For both groups, the scree and parallel analysis criteria (Lautenschlager, 1989) suggested a four factor solution. Oblique (Promax) solutions yielded the same factor structures as the orthogonal (Varimax) rotations, therefore the results of the orthogonal rotations are interpreted. For the honest group, factor 1 was interpreted as conscientiousness and emphasized working hard. The second factor was interpreted as agreeableness and emphasized lack of anger and grudges. The third factor was

also an agreeableness factor and emphasized easily getting along with others. The final factor was a second conscientiousness factor which emphasized avoidance of trouble. The four factors contained both biodata and personality items and were clearly interpretable as construct factors. In contrast to the factor results for the honest group, the four factor solution for the faking group yielded two methods factors and two construct factors. The first factor was a biodata methods factor. The second factor was interpreted as conscientiousness. The third factor was a personality methods factor. The fourth factor was interpreted as agreeableness. Thus, the empirically-derived factor structure was degraded by faking responses.

### **Criterion-Related Validity Analyses**

Table 2 presents the validities for the four rationally-derived scales, the four scales derived from factor analysis of the data provided by the honestly responding subjects, and the four scales derived by the factor analysis of the data provided by the faking subjects. The scale validities yield a consistent pattern. Except for the personality methods factor, the validities are at useful levels (.18 to .37) in the honest condition. In the faking condition, all validities are substantially lower than in the honest condition (-.09 to .15).

-----  
Table 2 about here  
-----

### **DISCUSSION**

Our results show that how a subject responds has implications for a variety of scale statistics. First, we confirmed that when subjects desire to do so, they can substantially improve their scores on both biodata and personality scales. This is consistent with a long line of research on faking (Hough et al., 1990; Kluger et al., 1991; Ones, Viswesvaran, & Korb, 1995). Given that faking can be situationally induced, we are puzzled with the attempts of many personality researchers to treat response bias as a personality trait. We agree that the tendency to respond favorably is correlated with stable personality traits (Ones, Viswesvaran, & Reiss, 1995). However, it is also clear that faking can be situationally induced, either by an experimenter or by the wishes of an applicant. We argue that this faking has substantial consequences for the use of non-cognitive measures with respect to reliability, construct validity, and criterion-related validity.

Faking serves to increase the internal consistency of scales. In honest responding, applicants truthfully report that their behavior is sometimes inconsistent. This truthful reporting hurts the homogeneity of the items. Faking respondents seek to report, with consistency, positive behaviors and thus their responses are more homogeneous. The high degree of response consistency, which we argue is characteristic of faking applicants, while improving internal consistency reliability, tends to destroy the construct validity of the scales as measured by multi-trait and multi-method analyses, and as indicated in the factor analyses. Thus an instrument, which measures multiple constructs for honest respondents, when completed by respondents who fake, may have a different factor structure. We speculate that respondent faking may be responsible for the sixth factor which is occasionally found in factor analyses of Big 5 measures (Cellar, Miller,

Doverspike, & Klawnsky, 1995; Schmit & Ryan, 1993).

Faking also decays the criterion-related validity of the scales. Scales which have useful levels of validity when administered to individuals responding honestly, have near zero validities when administered to respondents who fake. In Table 2, the average validity of the scales was .26 for the honest subjects and .04 for the faking subjects. The decay in the criterion-related validity is not a range restriction problem.

We believe that the poor construct and criterion-related validities for scales completed by applicants who are faking has two important implications for personnel selection. We describe each of these implications below.

First, these findings help clarify the results of several validity generalization studies of non-cognitive measures (Barrick & Mount, 1991; Ones, Viswesvaran, & Schmidt, 1993; Rothstein, et al., 1990; Tett et al., 1991). Most validity generalization studies of non-cognitive measures show some unexplained variance in the population validity distribution. We argue that some of this unexplained variance is attributable to differences across studies in the proportion of applicants who are faking. When a high percentage of the applicants are faking, the validity will be low or zero. When a small percentage of the applicants are faking, the validity will be higher.

It is informative to compare our validity coefficients with those found by Barrick and Mount (1991). These authors found a mean observed validity of .15 for conscientiousness and .05 for agreeableness. Thus, their average validities are much lower than the validity obtained for our honest subjects and are more consistent with the validities shown for our faking subjects or with the validities obtained when we combine the faking and honest subjects into one sample. Based on this observation, we speculate that the respondents in the typical validity study summarized by Barrick and Mount (1991) included a substantial number of people who were faking good.

Second, we argue that one can never be sure of the validity of a non-cognitive measure for a given sample. The criterion-related validity of non-cognitive measures will vary with the proportion of the applicants who are faking. If most of the applicants are faking, the criterion-related validity will be low or zero. Whereas, one does not know when applicants are faking, one can never be sure that non-cognitive measures are valid. This is a disquieting conclusion for those seeking to use non-cognitive measures in personnel screening. However, we suggest one set of indicators that would permit a probabilistic statement concerning the degree to which faking is a problem in any given sample. We found the analyses of internal consistency reliability, multi-trait-multi-method validities, factor analyses, and mean scores vary as a function of whether respondents are faking or not. We suggest that one can experimentally manipulate faking, as was done in this study, and run those analyses which are sensitive to faking. When operationally using the instrument, one can run the same analyses and compare the results to those obtained from the faking and honest samples. When the operational results are similar to the faking group, one might conclude that there is substantial faking present in the operational sample.

This study presents conclusions, which if true, present major problems for the application of personality and biodata measures in personnel selection. For this reason, it should be subject to

extensive review. Therefore, we solicited and assembled a list of criticisms. This information is presented as an appendix which is not included here due to space restrictions. The appendix is available from the authors.

## CONCLUSION

The present authors have used and advocated non-cognitive measures in applied settings and have published research showing useful levels of validity for non-cognitive measures. We were therefore quite surprised with the magnitude of the construct and criterion-related differences between faking and honest conditions. We are sympathetic to the concerns of researchers of non-cognitive measures and believe that the concerns are best addressed through replication and extension of this research. Replication using job applicants, an assortment of jobs which employ various non-cognitive and job performance measures, and which permit an analysis of individual differences in faking behavior would go far in extending our knowledge.

In summary, this study raises questions concerning the validity of non-cognitive measures when applicants fake good. We do not argue that non-cognitive measures will fail in all selection contexts. In many applicant pools, there may be relatively little faking. However it is clear that applicants can fake when they choose to do so. Our data demonstrate that when applicants choose to fake, validity will decay substantially.

## REFERENCES

- Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44:1-26.
- Cellar, D.F., Miller, M.L., Doverspike, D. D., & Klawnsky, J.D. (1995). A comparison of factor structures and criterion-related validity coefficients for personality measures based on the lexical and questionnaire five-factor models: The impact of scale transparency on factor structure and prediction. Manuscript submitted for publication.
- Christiansen, N.D., Goffin, R.D., Johnston, J.G. & Rothstein, M.G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47:847-860.
- 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. & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G.S. Stokes, M.D. Mumford, & W.A. Owens (eds). *Biodata Handbook: Theory, research and use of biographical information in selection and performance prediction*. Palo Alto: Consulting Psychologist Press.

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.

Lautenschlager, G.J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24:365-395.

Lautenschlager, G.J. (1994). Accuracy and faking of background data. In G.S. Stokes, M.D. Mumford, & W.A. Owens (eds). *Biodata Handbook: Theory, research and use of biographical information in selection and performance prediction*. Palo Alto: Consulting Psychologist Press.

Ones, D.S., Viswesvaran, C., & Korb, W.P. (1995, May). *Meta-analysis of fakability estimates: Between-subjects versus within-subjects designs*. Paper presented at the tenth annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1995, May). *Social desirability in personality testing: The red herring*. Paper presented at the tenth annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Ones, D.S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78:679-703.

Rothstein, H.R., Schmidt, F.L., Erwin, F.W., Owens, W.A. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, 75:175-184.

Schmit, M.J., & Ryan, A.M. (1993) The big five in personnel selection: factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78:966-974.

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

Tett, R.P., Jackson, D.N., Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44:703-742.



Table 1

Multi-trait-multi-method matrices for the four rationally-keyed scales.  
Presented separately for combined conditions, honest condition, and faking condition.

Combined Conditions N = 597 - 600							
				Agreeableness		Conscientiousness	
				Biodata	Personality	Biodata	Personality
Agreeableness - Biodata	.81	56.5	11.8	1	0.69	0.68	0.56
Agreeableness - Personality	.82	53.6	9.1		1	0.47	0.54
Conscientiousness - Biodata	.83	55.9	11.1			1	0.72
Conscientiousness - Personality	.83	54.4	9.8				1
Honest Condition N = 293							
Agreeableness - Biodata	.69	50	10	1	0.61	0.39	0.33
Agreeableness - Personality	.79	50	10		1	0.27	0.35
Conscientiousness - Biodata	.73	50	10			1	0.65
Conscientiousness - Personality	.78	50	10				1
Faking Condition N = 304 - 307							
Agreeableness - Biodata	.79	62.7	10.0	1	0.65	0.73	0.54
Agreeableness - Personality	.79	57.2	6.5		1	0.47	0.63
Conscientiousness - Biodata	.84	61.4	9.1			1	0.64
Conscientiousness - Personality	.80	58.6	7.5				1

Table 2

## Criterion Related Validity of Non-Cognitive Scales

Scales	Combined Conditions N=206-210	Honest Condition N=97	Faking Condition N=109-111
Rationally-Keyed Scales			
Agreeableness (biodata)	.08	.21	.12
Agreeableness (personality)	.14	.37	.00
Conscientiousness (biodata)	.11	.31	.06
Conscientiousness (personality)	.07	.24	-.02
Scales From Factor Analysis of Honest Respondents			
Conscientiousness (work hard)	.10	.28	.03
Agreeableness (little anger)	.12	.28	.12
Agreeableness (easy to get along with)	.12	.18	.15
Conscientiousness (avoid trouble)	.07	.29	-.09
Scales From Factor Analysis of Faking Respondents			
Biodata methods factor	.09	.33	.07
Conscientiousness	.08	.24	-.06
Personality methods factor	.01	.05	.00
Agreeableness	.15	.28	.13

Appendix To  
Douglas, E.F., McDaniel, M.A., and Snell, A.F. (1996)  
The Validity of Non-Cognitive Measures Decays When Applicants Fake  
A paper presented at the annual conference of the Academy of Management.  
Cincinnati. August.

The purpose of this appendix to Douglas, McDaniel and Snell (1996) is to address the additional issues that were not included in the paper due to length restrictions of the Academy of Management Proceedings (August, 1996). This paper also offers the opportunity to address criticisms that we have received concerning the paper based on an informal distribution of the paper prior to the conference. First, we will present an analysis indicating that one needs only a small percentage of applicants to fake in order to have the majority of applicants in the top hiring range to be fakers. Next we present a model of faking consistent with our data and related literature. Finally, we present the criticisms of our work and our rebuttal statements.

Fakers Rise to the Top: A Replication of Zickar, Rosse, and Levin (1996)

We used our data to attempt a replication of the findings of Zickar, Rosse, and Levin (1996). Zickar et al. conducted a Monte Carlo simulation in which they used item response theory to model faking on a non-cognitive measure. They demonstrated that one needed relatively few fakers in the sample for the top end of the distribution to contain a high percentage of fakers. Fakers at the high end of the score distribution are problematic because an employer typically hires applicants from the high end of the score distribution. Zickar et al. found that as the percentage of fakers in the sample increased, the percentage of fakers in the top end of the distribution increased although the validity of the test was not substantially moderated by the percentage of fakers in the sample.

In our replication, we conducted a Monte Carlo simulation using the data from this study. Subsets of subjects in the faking sample were randomly selected and added to the 97 honest subjects, so that the resulting sample had either 10, 15, 20, or 25% faking subjects. Validity coefficients for measures of agreeableness and conscientiousness were then calculated. Also the number of fakers in the top 10 individuals in each sample, rank ordered by descending agreeableness or conscientiousness scores, were counted. This was done 500 times for each of the 10, 15, 20 and 25% samples. The mean validities and mean numbers of fakers across 500 of these iterations are reported in Table A-1.

As seen from Table A-1, if 10% of the people fake, 5 to 6 of the top 10 subjects may be fakers and for these fakers the validity of the test is less than that for the honest subjects. The 25% faking sample is the worst scenario we examined. When 25% of the sample is faking, 9 of the top 10 subjects are fakers. Even though the validity of the full sample is about .20, the validity for these fakers is near zero.

Table A-1

Validity of Agreeableness and Conscientiousness  
with Varying Percentages of Faking Applicants

Percent Faking	Agreeableness		Conscientiousness	
	Mean Validity	# of fakers in top 10 respondents	Mean Validity	# of fakers in top 10 respondents
0%	.34	0	.30	0
10%	.27	5.8	.25	4.9
15%	.24	7.4	.23	7.1
20%	.22	8.4	.21	8.5
25%	.20	8.8	.19	8.9

Table Note: The agreeableness measure in this table is the sum of the agreeableness-biodata scale and the agreeableness-personality scale discussed in the *Measures* section of the Academy of Management paper. The conscientiousness measure in this table is the sum of the conscientiousness-biodata scale and the conscientiousness-personality scale discussed in the *Measures* section of the Academy of Management paper. Subsets of respondents in the faking sample were randomly selected and added to the 97 honest subjects, so that the resulting sample has either 10, 15, 20, or 25% faking subjects. The validity of the measures of agreeableness and conscientiousness for the prediction of job performance were then calculated. Also the number of fakers in the top 10 applicants of each sample, rank ordered by descending agreeableness or conscientiousness scores, were counted. This was done 500 times for each of 10, 15, 20, and 25% samples. The numbers reported in the tables are the mean validities and the mean numbers of fakers across these 500 iterations.

Our results and those of Zickar et al. indicate that the common finding that some measures of personality (e.g., conscientiousness) typically yield positive validities does not mean that there are no applicants faking. It appears that even with relatively small proportions of faking applicants, most of those in the hiring range are fakers.

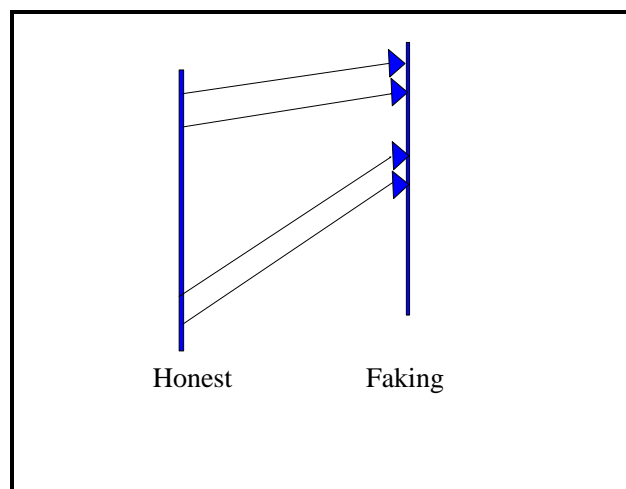
We note that our results are different from Zickar et al. in that in our findings the validities dropped with increasing proportions of fakers. No such drop in validity was seen in the Zickar et al. study. Michael Zickar ran some additional simulations (Personal communication, July, 1996) and found that the validity decay is a function of the variance of faking in the sample. The validity declines as the variance of the faking increases. Our attempts to resolve the discrepancy between the initial Zickar et al. analyses and our results have led us to develop a simple model of faking.

### A simple model of faking

Faking might be viewed as a variable that is added to one's test score. If all applicants fake equally, such as if no one fakes or if everyone fakes the same amount, faking is a constant and its variance would be zero. If one added this faking constant to every sample members' test score, the correlation between the test and other variables, such as job performance, would not change. However, if some applicants fake more than others, the faking variable would have variance. Some applicants' scores would increase more than other applicants. This would tend to change the rank order of the applicants and would change the magnitude of the correlation between the test and job performance. A faking model which is consistent with our data would need to explain how some applicants might fake more than others (i.e., how the faking variance would be non-zero). Here we seek to offer a simple model of faking that is consistent with our data. We suggest that faking is a function of three classes of variables: opportunity to fake, personal characteristics, and situational factors. We will describe each class of variables below.

Some applicants have more opportunity to fake than others. Individual differences in the opportunity to fake is a function of their true standing on the variable being measured. Those whose true behavior is substantially positive have little opportunity to fake good. Thus, those applicants who have never stolen from an employer do not have the opportunity to fake positively on the item "Have you ever stolen from an employer?" In contrast, those whose true behavior is very deviant have substantial opportunity to fake. An accomplished thief has the choice of answering honestly or lying to make a good impression. We illustrate this relationship between true trait scores and the opportunity to fake in Figure 1. As shown, those with a true low standing on the trait have more opportunity to fake than those with a true high standing on the trait, but the rank order of the respondents is the same in both honest and faking conditions.

Opportunity to fake may also be a function of other factors. For example, an applicant who is presently an employee of the company and whose work record is available to the personnel screener has less opportunity to fake than an outside applicant where little information is available to verify the applicant's claims. In addition, opportunity to fake may vary with the characteristics of the items. For example, Alliger, Lilenfield, and Mitchell (1995) found that subtle questions were less easily faked than questions where it was obvious what the measure was assessing. Items presented orally, such as in employment interviews, might be more or less fakable than items on paper-and-pencil measures. We suspect that the opportunity to fake in an interview might be more difficult because the interviewer could probe for additional information when faced



**Figure 1.** Applicants with low true scores on a trait have more opportunity to improve their scores through faking than do applicants with high true scores on a trait.

with a suspicious answer. However, there are a substantial number of coaching resources (e.g., books, courses) available to interviewees instructing them on ways to make a good impression in the interview. Thus, some may argue that it is easier to fake in an interview than on a non-cognitive paper-and-pencil test. Finally, the construct being measured may also influence the opportunity to fake. We believe that one has greater opportunity to fake on non-cognitive measures than cognitive measures. Thus, through faking, one might be able to improve her score on a measure of conscientiousness, but not on a vocabulary test.

Figure 1 would reflect the state of affairs if there are no personal or situational differences influencing the tendency to fake. Below, we speculate on personal and situational characteristics that would make Figure 1 unrealistic.

Personal characteristics may influence the tendency to fake. Some individuals' behavior is strongly influenced by personal or religious values (standards, tenets) that cause them to present themselves accurately even if it includes admitting to undesirable behavior and results in the loss of a valued outcome, such as a desired job. Other individuals will freely distort the facts to gain advantage. These individual differences will cause some people to fake more than others. Another personal characteristic is ability to fake. Some are likely to be better at faking than others. For example, those with knowledge of the construct being measured might be better able to fake than those who are unaware of the construct being measured.

Some of these individual differences may be related to the constructs that the non-cognitive measures seek to assess. For example, those whose conduct is heavily influenced by a moral code detailing right and wrong behavior may tend to score favorably on measures of conscientiousness. Other characteristics that influence faking, such as risk taking, may be unrelated to the construct the measure attempts to assess. Risk takers might fake more on a test of agreeableness for a customer service job and risk taking may be uncorrelated with agreeableness.

Finally, faking may be influenced by situational characteristics. A person whose job serves their needs (e.g., financial, emotional) may have few situational pressures encouraging them to fake when applying for a job. In contrast, a person who hates their present job and who cannot meet their financial responsibilities with their present job may have substantial situational pressures to fake good. For example, consider the situation of an individual who has been fired as a result of a layoff. If the person has substantial financial pressures (e.g., housing payments, car payments, children in college), he may be more likely to fake than most others.

Collectively, these factors are likely to cause individual differences in faking. When the factors cause substantial variance in faking, the validity should drop substantially when compared to a sample of honest respondents because the individual differences in faking will alter the rank order of the respondents relative to their rank order when responding honestly. If the factors cause little faking variance, the faking should cause little decay in validities because the rank order of respondents should not be altered substantially. Note that the average level of faking may be unimportant. If everyone fakes the same amount, even if it is a huge amount of faking, the validities of the measure should be the same as in the situation where everyone is honest.

Figure 2 graphically displays the effect that individual differences in personal and situational characteristics might have in influencing the extent to which respondents can capitalize on their opportunity to fake. These individual differences result in a change in the rank order of applicants when they respond to fake good relative to their rank order when responding honestly. These changes in rank order can explain the validity decline when applicants fake.

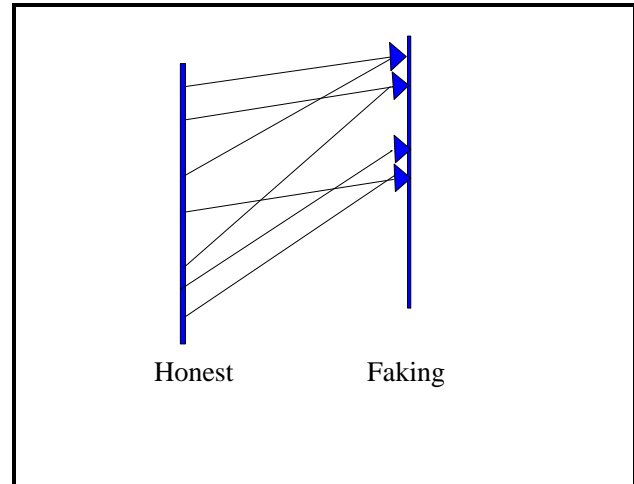


Figure 2. Individual differences in propensity to fake good alters the rank order of the applicants.

As faking variance increases (that is, as the individual differences in the degree of faking increases), the variance of the observed score on the non-cognitive measure may change. We suspect that the variance of the observed score will decrease with increasing faking variance. This may seem counter intuitive. If one adds variance to true score variance, one would normally expect an increase in observed variance. We suspect that the increased faking variance will reduce the observed variance because opportunity to fake is inversely-related to scores on the construct being measured. Those who are truly low on the non-cognitive construct being assessed have more opportunity to fake (i.e., more room for score improvement) and thus may engage in more faking. This will raise their observed scores so that they are closer to the observed scores of those high on the construct. Those high on the construct have little opportunity to fake and thus cannot improve their scores very much through faking. Thus, with substantial faking variance, those whose true scores on the construct is high and those whose true scores on the construct is low will both tend to have high observed scores. Thus, when people fake, the variance of their observed scores should be smaller than the variance of observed scores when the respondents are honest.

Our data support the hypothesis that high faking variance will result in lower observed score variance relative to groups with low faking variance. We assume that our faking group has higher faking variance than our honest group because the validities moved to zero and the honest responding instructions would cause the mean faking in the honest group to be near zero and nearly invariant. In the three of the four scales shown in Table 1, the variance of the observed scale scores is lower in the faking groups than in the honest group.

Wheeler, Hamill, and Tippins (1996) also presented data supporting the reduction of observed variance when applicants fake. A personality battery was administered in two time periods. In the first time period, the applicants received a warning not to fake. In the second time period, no warning was given. The scores during the second time period were substantially higher suggesting that the warning in time 1 was effective in reducing faking. The variance of the observed scores was smaller in time 2 than in time 1. We suggest that the faking variance caused this reduction in the observed variance.

### Critiques and rebuttals

Our results are generally perceived to be discouraging concerning the use of non-cognitive measures in personnel screening. The findings were surprising to us given that the authors have conducted previous research with non-cognitive measures and have been advocates of their use. Whereas we believe that our results should be examined closely, we have distributed our Academy of Management paper to various researchers in the field of personality testing in personnel selection. Many have graciously provided us with critiques of our paper. Below we present each critique and follow the critique with our attempted rebuttal.

*Criticism #1: Research has demonstrated that both personality measures and biodata measures can and do predict job performance (Asher & Sciarrino, 1974; Barrick & Mount, 1991; Hunter & Hunter, 1984; Reilly & Chao, 1982; Tett, Jackson & Rothstein, 1991). Therefore, the issue of faking is not relevant to the applied use of non-cognitive measures in personnel screening.*

We concur that a substantial body of literature yields positive validity coefficients for both personality and biodata measures. As seen from the results of our Monte Carlo simulation and the results of Zickar, et al., the fact that personality tests have positive validities for predicting job performance does not refute the hypothesis that faking decays validity and that fakers, for whom the test may have zero validity, are the majority of the applicants in the hiring range. One can obtain a large validity coefficient for a non-cognitive test for the full sample, but have a zero validity coefficient for those who are hired.

*Criticism #2: Applicants respond to questions with their true opinions. You instructed subjects to ignore their true opinions and respond in a favorable way. Thus, your results don't have relevance to actual applicant screening. Therefore, the paper is of little interest from a scientific perspective.*

We disagree with the reviewer's assertion that all people provide their "true opinions" when responding to a non-cognitive test in an employment screening context. We believe that it is reasonable that at least some applicants in some screening situations will fake their responses to look good. In a comprehensive review of the faking literature, Rynes (1993) concluded that faking is a common occurrence in applicant screening. The Wheeler et al. study also provided strong evidence that applicants will fake. Thus, we suggest that our results have substantial relevance to actual applicant screening. We believe that many will find this suggestion interesting from a scientific perspective.

Although it appears that some researchers believe that no one fakes in applicant screening settings, many large organizations do not use non-cognitive measures because they believe that such measures can be easily faked. For example, despite spending millions of dollars for the development of the Assessment of Background and Life Experiences (ABLE), a non-cognitive measure, the U.S. Department of Defense refuses to use the test for the screening and selection of military recruits. Additionally, a major industry consortium has dropped non-cognitive measures from its employment screening batteries due to decaying criterion-related validities over time attributed to the increased usage of coaching strategies by unionized workers (Personal communication to authors, August, 1995).



We suggest that faking may be a common occurrence particularly when it is facilitated by formal and informal applicant and employee groups. Such informal groups may consist of individuals who know each other through social networks, such as a group of friends who are all seeking employment with the same employer. This is a common occurrence in smaller "company" towns which contain only a few large employers. Formal groups would typically be labor unions or other labor organizations. Unions are the prime example of groups that seek strategies to improve scores on selection measures because unions strongly prefer seniority over formal testing programs for selection and promotion decisions. Thus, unions often pursue strategies to improve the test scores of their members. These formal and informal groups promote the use of strategies for faking. These strategies may produce individual differences in faking due both to individual differences in the opportunity to fake and due to individual differences in the ability to apply the strategies. These individual differences in degree of faking would likely alter the rank order of the applicants and attenuate the validity of non-cognitive measures. These strategies might also increase the number of applicants who can fake effectively with the result that most of the high scoring applicants will be fakers.

Several strategies may be used to fake. The theft of answer keys is sometimes attempted. The authors know of one unionized trucking company that was forced to stop using a standardized personality measure after most applicants began obtaining perfect scores. This incident was presumed to be related to the theft or illicit sale of an answer key. However, the most common strategy used to improve scores is through coaching. Coaching strategies can take at least three forms. First, respondents are instructed to think of an employee who is highly successful on the job and answer each question as they would see this person answering the questions. Second, respondents learn to emulate the responses of one who scored well on a previous administration of the test. Third, respondents receive explicit instructions on how to respond to various types of questions. Alliger, Lilenfield and Mitchell (1995) described a coaching program for integrity tests. Subjects were taught the mnemonic "PLACE" where each letter represented a rule to follow when answering questions. For example, "P" stood for protect. In responding to questions, respondents should indicate that they would not protect dishonest coworkers from punishment. Note that coaching can affect scores on cognitive tests, but typically, only through increasing the actual cognitive skills tapped by the tests. However, coaching can affect the score on a non-cognitive measure, such as a test of conscientiousness, without increasing the conscientiousness of the applicant.

*Criticism #3: Some discount the opinions of Rynes (1993). There is both large sample data (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990) and persuasive narrative reviews (Hogan, 1991) which indicate that applicants either do not fake to a large degree or they don't fake at all.*

It has been shown that faking can occur with the use of non-cognitive measures (Hough et al., 1990; Kluger, Reilly & Russell, 1991; Ones, Viswesvaran, & Korb, 1995). However, how often faking actually occurs in selection situations is still a debated question. A number of studies have provided evidence for the argument that in actual selection situations, faking either does not occur, or if it does occur, it has little effect on the validity of measures (Dunnette, McCartney, Carlson, & Kirchner, 1962; Abrahams, Neuman, & Githens, 1971; Hough et al., 1990; Hogan, 1991). We find this literature sparse and unconvincing. Due to its large samples, we consider the Hough et al. (1990) study to be the most persuasive article purporting that applicants do not

usually fake in applicant settings. Hough et al. demonstrated that applicants for the military service have scores similar to subjects instructed to respond honestly on a battery of non-cognitive tests. We believe that the Hough et al. study lacks generalizability to most applicant screening situations. Applicants to the military are typically about 18 years old and have had little experience in the job application setting. Thus, they may not know the potential value of misrepresenting oneself in seeking employment. We speculate that a group of downsized 40 year olds with hungry children and late mortgage payments would respond with substantially more faking.

The Hough et al. study might also be less compelling when one considers our results and those of Zickar et al. which suggest that one needs relatively few fakers for the fakers to rise to the top of a hiring list sorted by descending scores. Consider the possibility that in the Hough et al. data only one or two percent of the applicants were faking. These relatively few numbers of fakers would not cause the means between the applicants and the honest responders to be very different. However, one might find these fakers to be at the top of the hiring list and thus result in poor selection decisions.

Our reading of the literature leads us to believe that little is known concerning how often people fake. We do not know if it is a big problem or a small problem. We do not know the extent to which the degree of faking varies from one applicant situation to another or what might cause the variation. We do know that applicants can fake if they choose to do so. Furthermore, the present data set indicates that when applicants fake, the validity suffers. Therefore, in contrast to Ones, Viswesvaran, and Reiss (in press) who dismissed the role of social desirability in personnel selection as a "red herring", we suggest that the issue of faking in non-cognitive measures is a topic that warrants continued debate. We believe that much more research is needed to determine the actual amount of faking in applicant settings and conditions under which the effects of faking can be minimized.

*Criticism #4: The selection scenario in this study did not realistically simulate the job selection context. The pressure to "distort" is much more subtle for real applicants.*

We concur that our study may not realistically simulate the job selection context, but that was not its purpose. The study was designed to manipulate faking motivation and examine the influence of faking on validity. It achieved this goal. We see four possible interpretations for the concept of "subtle faking."

First, subtle pressure might result in variable degrees of faking. For example, many applicants may be less motivated to fake than those subjects in our faking condition. When there is little variance in faking such as in applicant pools where no one fakes or when all subjects fake equally, we do not expect any decay in validity. However, when there is faking variance, the rank order of the applicants will vary from the rank order obtained when all applicants respond honestly or when all applicants fake to the same degree. Under most such circumstances, the validity should decline.

Second, subtle pressure might mean that those who fake subtly behave in a manner that is different from those who fake with abandon. For example, personality researchers could fake

differently from typical applicants such that researchers could avoid endorsing items on social desirability scales. Our data shows validity decays when people fake on substantive personality scales. Although we agree that applicants can employ various faking strategies, there is no empirical evidence that different faking strategies would alter our conclusion that validity decays when respondents fake. As long as there is variance in faking, we suspect that there will be validity decay regardless of what faking strategy or strategies may be employed by the applicants. We agree that research examining various faking strategies is warranted.

Third, it has been shown that subjects are able to fake differently for different jobs and this may have different effects on validity. Kroger and Turnbull (1975) found that when subjects possess an accurate conception of the role to be simulated, they are able to fake personality tests by enacting a specific social role, rather than by responding in terms of the personality constructs. This method of role faking, according to their results, cannot be detected by validity scales. Additionally, results of a study by Mahar, Cologon and Duck (1995) supported the suggestion that respondents use a stereotype of workers in the relevant occupation when faking personality tests for vocational advantage. The studies suggest that applicants are not naive and not fully truthful. Whether motivation to fake in specific jobs causes faking strategies and validity patterns to differ is an empirical question that deserves further attention.

Fourth, subtle pressure might refer to differences in testing motivation or affect associated with faking. Our subjects faked in response to instructions to fake and were motivated solely to cooperate with the researchers. There is likely to be little affect associated with their faking behaviors. Applicants complete non-cognitive measures with the motivation to obtain a desired job. Applicants may have more affect associated with their faking behavior. For example, they may fear being accused of lying. We agree that our subjects may have had less emotional arousal than the typical applicant. We agree that our results need replication with real applicants.

In concluding our rebuttal to this criticism, we note that subtle faking might cause even more individual differences in faking than in our study. In our faking instructions, we encouraged the respondents to fake. Thus, we likely reduced individual differences in the willingness to fake and thus might have reduced the faking variance due to personal characteristics. In an actual applicant setting, some applicants may be more willing to fake than others. These individual differences may result in larger faking variance than in our study.

*Criticism #5: It is not feasible to argue the existence of stable personality traits that correlate with the tendency to respond favorably AND the existence of situationally-induced faking.*

The existence of one does not rule out the existence of the other. Individual differences in these stable traits will determine some proportion of the faking that occurs. However, when certain situations present themselves, such as motivation to fake to obtain a job, situationally-induced faking will be present and contribute to the degree of faking that occurs. One can be anxious because one is neurotic. One can also be anxious because someone is pointing a gun at them. Likewise, one can inflate her score on a measure of conscientiousness because she is traited to do so and also because she needs a job because her family is starving. When one partials social desirability variance from a personality test, one is partialing variance that may largely result from

the stable personality traits. The residual test score resulting from this partialing may still contain variance attributable to situationally-induced faking.

*Criticism #6: This study used locally-developed scales. Thus, the results should be questioned.*

We agree that published personality scales exist for conscientiousness and agreeableness, but note that no published biodata scales are available for conscientiousness and agreeableness. Thus, at least the locally-developed biodata scales were necessary. We suggest that, for the honest respondents, the multi-trait, multi-method analyses and the factor analyses indicated that both the rational scales and the factor-analysis-derived scales have construct validity. The internal consistency reliabilities of the scales were also acceptable. Thus, we believe that there is nothing peculiar about these scales which would cause one to discount the validity results in this study. We agree that our scales would have greater credibility if they were correlated with more established measures. We also agree that one could have greater confidence in our results if they were replicated with published, well-researched scales.

*Criticism #7: Conscientiousness and agreeableness measures contain items with high correlations with socially desirable responding. Thus, the researchers "stacked the deck" against non-cognitive measures by picking these two measures.*

We chose the constructs of agreeableness and conscientiousness because they have the best record for the prediction of job performance. Tett, Jackson and Rothstein (1991) found useful levels of validity for measures of both agreeableness and conscientiousness. Barrick and Mount (1991) also found useful levels of validity for conscientiousness. Thus we suggest that the constructs assessed are reasonable choices for a test battery designed to predict job performance.

*Criticism #8: This study used the same performance appraisal rating form for widely different jobs and for employees in widely different organizations. Therefore, the results are suspect and should be ignored.*

Our criterion was an adaption of a performance appraisal form used successfully for decades in the validation of the General Aptitude Test Battery (U.S. Department of Labor, 1977). Gandy and colleagues (Gandy, Dye, & MacLane, 1994; Gandy, Outerbridge, Sharf, & Dye, 1989) used the performance appraisal form developed for the General Aptitude Test Battery validation studies along with performance appraisal forms that were developed to be content specific to the jobs under study. They found comparable validities for the two rating criteria. We believe that our criterion measure provides a useful rank ordering of the individuals with respect to job performance. Furthermore, if problems are caused by the use of this rating instrument, the problems would affect both the honest and faking conditions. The finding that we could predict job performance in the honest condition demonstrated that the criterion was capturing reliable variance related to the predictors.

*Criticism #9: Faking has the likely effect of adding a constant to everyone's score. Therefore, the rank ordering of the individuals would not change and thus the validity would not change.*

Our data show that the validity of the non-cognitive measures substantially decrease when subjects fake and are substantially higher when subjects are instructed to respond honestly. Therefore, the rank order of the subjects in the honest and faking conditions cannot be the same. Our between-groups design does not let us demonstrate the changing rank order for any given individual. A within-subjects design would permit us to demonstrate that the rank order is not the same and we are presently collecting data on this issue. Still, the present data clearly show the rank order is not the same because the validities are not the same.

*Criticism 10: There is a substantial body of literature indicating that correcting for social desirability does not affect the validity of the scales. Therefore, the logical conclusion is that faking does not hurt validity. We believe that you dismiss that literature too readily.*

We concur that controlling for social desirability does not improve the validity of personality measures. However, as seen in our study, the validity does decay when respondents fake. One conclusion that might be reached is that social desirability scales do not capture all the variance associated with faking. This issue can be addressed empirically by including social desirability scales in a study similar to ours where some subjects fake and others respond honestly. One could pool the subjects from both conditions and determine the extent to which the social desirability scales can distinguish the fakers from the honest respondents.

We also note that the validity for the full sample, after being corrected for social desirability, may not adequately summarize the usefulness of the test. As seen in our data, one can obtain a large validity coefficient for a non-cognitive test for the full sample, but have a zero validity coefficient for those who are hired.

*Criticism #11. The authors conclude that one may never be sure of the validity of a non-cognitive measure for a given sample. This is an ugly revival of the long-discredited situational-specificity hypothesis. For biodata, the Rothstein, Schmidt, Erwin, and Owens (1990) meta-analysis refutes your situational specificity hypothesis. For conscientiousness, the Barrick and Mount (1991) meta-analysis refutes your situational specificity hypothesis. For integrity, the Ones, Viswesavaran, and Schmidt (1993) meta-analysis refutes your situational specificity hypothesis.*

In the Academy of Management proceedings paper, we stated that one can never be sure of the validity of a non-cognitive measure for a given sample because the criterion-related validity of the measure will be moderated by the percent of the respondents who are fakers and the number of fakers is unknown. After further consideration, we believe that the validity decay is primarily a function of the variance of faking. Specifically, the extent of faking does not alter the validity if everyone fakes the same amount. There must be individual differences in faking across respondents (i.e., faking variance) for the faking to alter the rank order of the respondents and thus alter the validity. Our Monte Carlo simulation finding showing that validity decays with increasing proportions of fakers is probably due to the fact that increasing the percent of fakers in the sample (from 0% to 25%) increases the variance of faking.

The situational-specificity hypothesis argues that the validity of a test will vary from situation to situation due to unknown moderators. Advocates of the situational-specificity hypothesis argue

that validity studies should be conducted in each setting where a test is used. We do not believe that our results have much in common with the situational-specificity hypothesis for two reasons. First, we contend that the variable that moderates the validity of non-cognitive measures is the extent to which there is variability in faking. Thus, the moderator is not unknown. Second, we do not advocate the conduct of separate validity studies in each setting. Such validity studies will not be informative because the validity coefficient for the full sample gives no indication of the validity of the test for the fakers who are likely to be the majority of those in the hiring range. Thus, we do not believe that our conclusions embody the major tenets of the situational-specificity hypothesis.

The reviewer's citations concerning various meta-analyses refuting our alleged situational-specificity hypothesis are misguided. Apparently the reviewer is asserting that if tests show validity generalization then faking is not a major problem in the use of non-cognitive tests for personnel selection. We agree that non-cognitive tests show validity generalization in that most of the validities are positive. However, we have shown that even when the validity for the full validation sample is positive and large, the validity for the fakers, who may be the majority of those hired, may well be zero.

*Criticism #12. The authors have made serious errors in their multi-trait multi-method analyses by failing to consider the effects of the differences in reliability of the scales between the faking and the honest groups.*

Table A-2 presents the information from Table 1 after correcting the correlations for measurement error in both variables. The results remain supportive of the conclusion that the construct validity of the measures, as evidenced by multi-trait multi-method analyses, is substantially better in the honest condition than in the faking condition.

*Criticism #13. The reliability differences for the scales across conditions is not large.*

We concur that although the scale reliabilities are higher in the faking condition than in the honest condition, the differences are not large.

*Criticism #14. The title of the paper is misleading. You did not have applicants in your sample.*

We concur that we did not have applicants in our sample. We agree that the replication of these results with actual applicants is warranted.

*Criticism #15. Your paper is written very forcefully. It could be understood if this study were conducted and interpreted by someone who is not familiar with meta-analysis. It is shocking that the authors, one with a strong background in meta-analysis, would be so confident about a small sample study. The authors should have known better. These results are probably due to sampling error.*

Table A-2

Results from Table 1 where the correlation matrices are corrected for measurement error in the scales

Combined Conditions N = 597 - 600							
	Alpha	Mean	$\sigma$	Agreeableness		Conscientiousness	
				Biodata	Personality	Biodata	Personality
Agreeableness - Biodata	.81	56.5	11.8	1	0.84	0.83	0.67
Agreeableness - Personality	.82	53.6	9.1		1	0.57	0.65
Conscientiousness - Biodata	.83	55.9	11.1			1	0.87
Conscientiousness - Personality	.83	54.4	9.8				1
Honest Condition N = 293							
Agreeableness - Biodata	.69	50	10	1	0.82	0.55	0.45
Agreeableness - Personality	.79	50	10		1	0.36	0.45
Conscientiousness - Biodata	.73	50	10			1	0.87
Conscientiousness - Personality	.78	50	10				1
Faking Condition N = 304 - 307							
Agreeableness - Biodata	.79	62.7	10.0	1	0.84	0.89	0.68
Agreeableness - Personality	.79	57.2	6.5		1	0.57	0.80
Conscientiousness - Biodata	.84	61.4	9.1			1	0.78
Conscientiousness - Personality	.80	58.6	7.5				1

We believe that researchers have paid insufficient attention to the effect of faking on the validity of non-cognitive measures. Strongly worded arguments are an effective way of bringing little researched issues to the attention of a wider audience. At a minimum, we have engaged many personality researchers in a debate of our findings. By winning the best paper award from the Academy of Management Human Resource Division, we believe we have been successful in bringing this issue to the attention of a wider audience. Whether the results are due to sampling error is an issue that is best resolved through replication.

*Criticism #16. Consider an analogy. Cognitive ability is an individual difference variable. Now imagine that you tell a group of subjects to mark on purpose the wrong answers on the test. You can thus situationally induce everyone to score lower than their real intelligence. Does this mean that cognitive ability is not an individual difference trait? I hope by now you are thinking "but nobody applying for a job would try to get the answers wrong on an ability test." And that is my*

*point. In order to carry through your argument with faking on non-cognitive measures, you need applicant data indicating what applicants do.*

We are confused by your analogy. We do not state that non-cognitive measures fail to assess stable individual differences. We do suggest that when there is variance in faking, measures of non-cognitive traits become less valid. We concur that replications with actual applicants is warranted.

*Criticism #17. Applicants are more likely to fake than employees participating in concurrent validity studies. Yet validity coefficients for non-cognitive measures in applicant samples are often substantial. If your conclusions are correct, the validity of non-cognitive measures in applicant samples should be zero.*

The validity decay is a function of the faking variance. If an applicant pool has very large faking variance (i.e., some people fake more than others), the validities may be near zero. If the applicant pool has zero faking variance, the validities will be unaffected by faking. Our data suggest that samples can yield moderate validities for the full sample yet zero validity for the subsample of fakers. We do not believe that the validity for typical applicant samples is zero.

We are unsure of the usefulness of the distinction between applicant and concurrent samples as it pertains to faking. We agree that applicants may be more motivated to fake than current employees. However, faking motivation is not the sole distinction between applicant and concurrent samples. In our consulting experience, we have often had difficulty with concurrent samples when the employees are participating against their will or are participating in a careless manner. Concurrent samples may have more participants responding randomly or carelessly than applicant samples. Such careless and random responding typically attenuates validity. In brief, there are more differences between applicant and concurrent samples than the motivation to fake. Thus, it is not clear that comparisons of applicant and concurrent samples shed light on the issue of faking.

*Criticism #18. Ones, Viswesavaran, and Schmidt (1993) found that the validity of integrity tests was invariant for job performance and thus no moderators can be operating. Therefore, for integrity tests, faking cannot moderate the validity.*

If the variance of faking was positive and constant across studies in the Ones et al. (1993) study, one could achieve validity invariance but still have a validity of zero for those respondents who were faking. However, we find it unlikely that variance of faking would be constant across so many studies. Another possibility is that there is zero faking variance on integrity tests because characteristics of the tests make them immune to faking. However, Alliger, et al. found that one could be trained to fake successfully on one test of integrity. Perhaps a replication of our study with integrity measures would shed light on the issue.

*Criticism #19: The difference in validities between the honest and faking group is due to range restriction in the faking group. Therefore, if you use a non-cognitive measure that is sensitive to differences at the favorable end of the trait being measured, faking will not cause range restriction, and faking will not reduce the validity of the non-cognitive measure.*



Range restriction, in the context of personnel selection, typically results from the selection of a sample based on scores on the predictor. For example, the correlation between a test and job performance can only be calculated on those individuals who scored high enough on the test to be hired. Neither the faking nor the honest samples were selected based on their scores on any predictor. The faking sample does have a smaller variance on the tests than the honest group and some might refer to that smaller variance as range restriction.

Earlier we suggested that an increase in faking variance serves to decrease the observed variance. However, it is not solely reduced observed variance that causes the decline in validity with faking. If one applied a range restriction correction to the faking group validity coefficients using the honest group predictor variance as the population variance, the adjusted faking group validities would still be low. Thus, it is not solely reduced observed variance that causes the validity decline in our data. The validity loss stems from a change in the rank ordering of the respondents caused by some applicants faking more than others.

Finally, we concur that faking can result in less observed variance and that the reduced variance can attenuate the validities. Consider the extreme example where applicants fake to the extent that all applicants have the highest score. For that circumstance, the test would have zero variance and the validity would be zero.

### Conclusion

Our Academy of Management paper has generated considerable controversy even before it has been presented. Thus, we have succeeded in our goal of raising interest in the effects of faking on the validity of non-cognitive measures. We have addressed all criticisms although our position is stronger on some than others. Frankly, we would like to be proven wrong because our conclusions are so potentially discouraging for the use of non-cognitive measures in personnel selection. We hope that additional research will discredit our findings.

References for Articles Cited in the Appendix but Not in the Primary Paper

- Alliger, G.M., Lilenfield, S.O., & Mitchell, K.E. (1995). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science*, 7, 32-39.
- Asher, J.J. & Sciarrino, J.A. (1974). Realistic work sample tests: a review. *Personnel Psychology*, 27, 519-533.
- Gandy, J.A., Dye, D.A., & MacLane, C.N. (1994). Federal government selection: The Individual Achievement record. In G.S. Stokes, M.D. Mumford, and W.A. Owens (Eds.) *Biodata handbook*. Palo Alto, CA: Consulting Psychologist Press.
- Gandy, J.A., Outerbridge, A.N., Sharf, J.C., & Dye, D.A. (1989, November). *Development and initial validation of the Individual Achievement Record*. Washington, DC: U.S. Office of Personnel Management.
- Hogan, R. (1991). Personality and personality measurement. In M.D. Dunnette & L.M. Hough (Eds.) *Handbook of industrial and organizational psychology* (Vol. 2, pp. 327-396). Palo Alto, CA: Consulting Psychologists Press.
- Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Kroger, R.O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology*, 43, 48-55.
- Mahar, D., Cologon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, 18, 605-609.
- Ones, D.S., Viswesvaran, C., & Reiss, A. (in press). The role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*.
- Reilly, R.R., & Chao, G.T. (1982). Validity and fairness of some alternative selection procedures. *Personnel Psychology*, 35, 1-62.
- Rynes, S.L. (1993) Who's selecting whom? Effects of selection practices on applicant attitudes and behavior. In N. Schmitt & W.C. Borman (Eds.), *Personnel selection in organizations*. San Francisco: Jossey Bass Publishers.
- Wheeler, J.K., Hamill, L.S., Tippins, N.T. (1996). *Warnings against candidate misrepresentations: Do they work?* Paper presented at the Eleventh Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

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