Measuring Faking Propensity

Michael D. Biderman
University of Tennessee at Chattanooga

Nhung T. Nguyen
Towson University

Authors' Note: Correspondence regarding this article should be sent to Michael Biderman, Department of Psychology / 2803, U.T. Chattanooga, 615 McCallie Ave., Chattanooga, TN 37403. Tel.: (423) 425-4268. Email: Michael-Biderman@utc.edu

Paper presented at the 24rd Annual Conference of The Society for Industrial and Organizational Psychology, New Orleans, LA, 2009.

Poster

TITLE

Measuring faking propensity

ABSTRACT

The utility of a measuring faking propensity as a method factor using single condition data was investigated. Faking propensity so measured demonstrated convergent validity with difference score measures of faking and moderate convergent validity with social desirability measures. Faking propensity was unrelated to cognitive ability, indicating discriminant validity.

PRESS PARAGRAPH

Whereas the fakability of Big Five personality questionnaires has been well documented, an accurate measure of faking by applicants has yet to be established. In this study, the utility of a method factor estimated within the context of a confirmatory factor analysis model was investigated. Faking propensity as measured by the method factor demonstrated convergent validity with traditional difference score measures of faking and moderate convergent validity with social desirability measures. Faking propensity was unrelated to cognitive ability, indicating discriminant validity. The potential of measuring faking as a method factor in operational settings of personnel selection is discussed.

The resurgence of the use of personality tests in employee selection has generated concomitant interest in the faking of those tests. There seems to be agreement among researchers on a general definition of faking as a systematic distortion in responses to personality questionnaire items causing them to differ from the responses that would have been made based on item content alone (Hough, 1998; Rosse, Stecher, Miller, & Levin, 1998; Zickar & Drasgow, 1996). On the other hand, there is less agreement on the measurement of faking. Three methods dominate the applicant faking literature. In this study, a fourth method considered since 1993 but not well investigated will be validated.

The most important reason that accurate measurement of faking is needed is that faking may result in hiring the wrong persons. For example, Mueller-Hanson, Heggestad, and Thornton (2003) found that faking changed the rank order of applicants such that for small selection ratios larger prediction errors were associated with those scoring high in incentive conditions. This indicates that some of those hired when selection ratios were small were not truly high on the personality characteristics of interest but faked to get to that level. This is a potential problem because hired fakers might perform below the expected performance level compared to their honest co-workers (Mueller-Hanson et al., 2003). A second reason for the need for accurate measurement is that faking may affect the construct validity of selection tests, limiting understanding of their true relationships to performance.

The first measure of faking is based on the premise that faking is the result of a disposition to exaggerate positive characteristics. This tendency was labeled socially desirable responding, and measures were developed to assess the tendency. Scores on these social desirability scales were assumed to measure the extent to which applicants were faking personality tests administered along with the social desirability scales. The most popular of such scales is that developed by Paulhus (1984, 1991), the Balanced Inventory of Desirable Responding (BIDR). The BIDR contains two subscales, one measuring Self Deception, a tendency to self-promote of which respondents are unaware, and Impression Management, a tendency to engage in calculated exaggeration of which respondents are aware.

A positive aspect of measuring faking using social desirability scales is that the measure of faking can be obtained in a single applicant setting, with the social desirability scale administered as part of the selection battery. The negative aspects are that their use has been of little benefit to research on faking. Research in the 1990s found no evidence that controlling for social desirability improved criterion related validity of personality tests or restored applicants' honest scores on such tests (e.g., Ellingson, Sackett, & Hough, 1999; Ones & Viswesvaran, 1998).

A second method of measuring faking has utilized differences in test scores between conditions conducive to faking and conditions in which respondents were believed to respond honestly. In between-subjects designs, mean differences between groups under the two instructional conditions have been used (e.g.,Viswesvaran & Ones, 1999). In within-subjects designs, difference scores for individual respondents have been computed and used to investigate the relationship of amount of faking to other factors. Notable among studies using such designs are those by McFarland and Ryan (2000, 2006) who administered personality questionnaires to respondents under instructions to respond honestly and again under instructions to fake. One important result of these studies is the finding that the ability to fake cut across personality scales suggesting that faking is a general characteristic.

If there is a gold standard measure of faking, it is probably difference scores computed from within-subjects designs. These scores have obvious face validity and require no cross-scale inferences such as those involved with the use of social desirability measures. They measure distortion of the actual personality dimensions that will ultimately be employed for selection. The primary difficulty with difference scores is their requirement of an honest-response condition. This severely limits the usefulness of the within-subjects design in applicant settings. The limitation of the design in turn limits the utility of difference score measures of faking. Clearly, the most useful measure of faking will be one that 1) involves only the questionnaires on which selection will be based and 2) can be applied to data of a single condition without the requirement of an honest response condition.

A third set of faking measures has been derived from application of Item Response Theory (IRT). For example, Zicker and Robie (1999) measured faking as theta shift between honest and faking conditions. Although promising, the IRT based methods suffer from lack of familiarity of researchers and practitioners with the methods.

A fourth way of measuring faking, one that has the potential to be used in single conditions and to involve only the personality questionnaires used for selection, is based on a representation of faking by a factor analogous to a method factor in multitrait-multimethod research (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Evidence for the presence of such a factor was found in a seminal study by Schmit and Ryan (1993). Their study was followed by Cellar, Miller, Doverspike, & Klawsky, 1996, who used confirmatory factor analysis to estimate a factor measuring faking. Cellar et al. found that adding a faking factor significantly improved goodness-of-fit to two personality questionnaires. Although there have been a few other studies in which faking has been represented by a method factor (Biderman & Nguyen, 2004; Jo, 1997, 2000; Klehe, Kleinmann, Hartstein, Melchers, & König, 2008), amazingly in the 15 years since the Schmit and Ryan (1993) study, the study of Cellar et al. (1996) is the only published article using a common factor definition of faking in a selection situation.

Figure 1 presents the model investigated by Cellar et al. (1996) as applied to Big Five personality data. Each Big Five dimension is represented by its own factor. The sixth factor is one that may represent faking. This model is analogous to Model 3A presented by Podsakoff et al. (2003) in their review of common method biases. Thus, whether or not the sixth factor represents method bias or faking is not determined by the model but by the conditions in which the model is applied. For that reason, it is labeled M/F in Figure 1.

Our goal in the present research was to examine the utility of the model of Figure 1 to measure faking from the data of a single condition as might be found in a selection situation. Even though the goal was to assess efficacy from one-condition data, a two-condition within-subjects design was employed in the belief that the two-condition data would provide the necessary perspective from which to evaluate the application of the model to one-condition data. In the design, participants were instructed to respond honestly in one condition and given an incentive to fake in the other. A generalization of the model of Figure 1 was applied to the two-condition data, with a method factor representing the honest condition and a separate factor representing the incentive condition as shown in Figure 2. The efficacy of the model of Figure 1 was assessed by applying it to only the data of the incentive condition and then comparing the results to those from the two-condition data.

Incentives rather than instructions to fake were used because incentives best approximated actual job application conditions. We believe that in such incentive-driven conditions differences in response distortion represent differences in faking propensity (e.g., Griffith, Chmielowski, Snell, Frei, & McDaniel, 2000). On the other hand distortion in conditions in which participants are instructed to fake irrespective of their motivation to gain an incentive is primarily a result of faking ability. For that reason, the factor indicated by incentive condition variables in Figure 2 is labeled FP. The factor indicated by the honest condition variables is labeled M since it was expected that individual differences not accounted for by the Big Five factors in the honest condition would be due to unspecified method effects.

Although three measures of faking have been used for much research, there have been few attempts to assess the convergent validity of the different measures. McFarland and Ryan (2006) found very small correlations between difference scores and the BIDR scales. Since they employed instructions to fake it is not certain whether it should be expected that the faking they observed would be correlated with social desirability. We are aware of no attempts to correlate social desirability with the method factor measure examined here. For that reason, the BIDR was administered and correlations of the faking factor with both difference scores and with the SD and IM scales were computed. We expected positive correlations of faking propensity as assessed by the faking factor with both difference scores and the SD and IM scales.

One of the major reasons for using personality tests has been the existence of adverse impact when cognitive ability tests are used in selection. Although there is evidence that faking ability following instructions to fake is correlated with cognitive ability (Biderman & Nguyen, 2004; Kuncel & Borneman, 2007) there has been little research on the relationship of faking propensity to cognitive ability. For that reason, the Wonderlic Personnel Test (WPT; Wonderlic, 1999) was administered to all respondents. We expected negligible correlations of faking propensity with cognitive ability.

METHOD

Participants.

Participants were 202 undergraduates enrolled at a mid-sized southeastern university in the USA. Respondents participated for extra credit.

Measures.

Wonderlic Personnel Test. The Wonderlic Form II was administered to 113 participants, and Form V to 89. Alpha reliability estimates were .80 and .78 for the two forms respectively.

Big Five dimensions. The 50-item Big Five questionnaire from the IPIP web site (Goldberg, 1999) was administered. Participants responded to each item using a seven-point scale, with one end of the response scale labeled “1=Completely Inaccurate” and the other end labeled “7=Completely Accurate”.

Social Desirability. The BIDR (Paulhus, 1991) was administered, and scored for SD and IM. Paulhus (1991) recommended counting the number of extremely positively directed responses - 6 or 7 from positively worded items, 1 or 2 from negatively worded items. Since that method ignores differences between persons who responded within the dichotomous categories, the SD and IM scores were instead computed as summated scores in the fashion recommended for such scores (e.g., Spector, 1992).

Difference scores. The mean of responses to all Big Five items from the honest condition was subtracted from the corresponding mean from the incentive condition to create a difference score measure of faking for each participant.

Method factors. For the two-condition data, confirmatory factor analyses were conducted applying variations of the model in Figure 2.[1] Specifically, four models were applied. In Model 1, all latent variables shown in Figure 2 were estimated. In Model 2, latent variable FP was omitted. In Model 3, M was omitted, and in Model 4, both FP and M were omitted. Omitting specific latent variables allowed their contribution to Model 1 goodness-of-fit to be evaluated using chi-square goodness-of-fit tests.

For one-condition data, two CFAs based on Figure 1 were applied to only the incentive condition data. In Model 1, all latent variables shown in Figure 1 were estimated. In Model 2, FP was omitted allowing evaluation of the contribution of FP to the incentive condition data.

Models were applied to individual item and to two-item parcel data. The two types of indicators were employed to insure that conclusions would be robust with respect to choice of indicator. All models were estimated with the method of maximum likelihood using MPlus V5.1 (Müthen & Müthen, 1998-2007).

Factor scores. Factor scores for the method factors were computed and imported into SPSS for analyses to establish convergent and discriminant validity and to assess correlations with variables external to the CFA model.

Procedure.

A within-subjects design was employed in which questionnaires were administered twice. In the first condition respondents were instructed to respond honestly. In the second condition, respondents were told that based on their responses to the second section of the questionnaire, "the twenty participants who would make the best candidates for employment will be entered into a drawing to receive one of four $50 Cash awards." Participants then filled out a prize drawing form. The phrase "Prize Drawing Questionnaire" was displayed at the top of all subsequent pages of the questionnaire. In order to insure that participants had no reason to fake in the honest condition, the incentive condition was last for all participants.

The WPT was administered first. Then the Big Five and the BIDR questionnaires were administered in each condition. Three other questionnaires were administered but not analyzed for the present report. Only the BIDR data from the honest condition were analyzed here.

RESULTS

Tables of correlations of individual indicators are too large to present here. To permit comparison with other research involving the same personality questionnaires correlations of scale scores from the two conditions are presented in Table 1. Inspection of the table indicates that the Big Five scales were moderately positively correlated in the Honest condition with mean correlation equal to .19, and more highly correlated in the incentive condition, with mean correlation of .30. It has been common to find Big Five scale scores more positively correlated in faking conditions (e.g., Douglas, McDaniel, & Snell, 1996) and this supports the efficacy of the incentive manipulation used here. Faking effect sizes were .12, .12, .29, .42, and .14 for extraversion, agreeableness, conscientiousness, stability, and openness respectively also supporting the expectation that the incentive condition was subject to a greater amount of faking than the honest condition.

Goodness-of-fit.

Two-condition models. Table 2 presents goodness-of-fit measures for the nested models based on the model of Figure 2. The first characteristic that is apparent in the table is that when items were indicators none of the goodness-of-fit statistics reached the commonly recommended thresholds of acceptability. On the other hand, when two-item parcels were indicators, goodness-of-fit statistics were close to acceptable values. Since there is accumulating evidence that goodness-of-fit is positively related to the coarseness of the indicators (e.g., Kenny & McCoach, 2003), we treated the poor values from the models applied to individual indicators as reflecting idiosyncratic, item-specific deviations rather than systemic failures of the model and continued to interpret the results based on individual items as well as those based on parcels.

A second result from Table 2 is that omitting either method factor resulted in a significant decrement to goodness-of-fit. This indicates that in both the honest and incentive conditions, these factors represent variance beyond that due simply to the Big Five factors. The third result notable in the table is that the decrement to goodness-of-fit was greater when FP was omitted than when M was omitted.. This suggests that the common variability in the incentive condition was greater than that in the honest condition.

One-condition models. Table 3 present goodness-of-fit statistics from application of the model of Figure 1 to the incentive condition data. As shown in Table 2, the statistics were closer to acceptable thresholds for the model applied to two-item parcels. Moreover, goodness-of-fit was significantly worse when FP was omitted.

Parameter Estimates.

Mean loadings are presented in Table 4. For the model applied to two-condition data, all mean loadings of items and parcels on the Big Five factors were positive with means from the incentive condition about equal to those in the honest condition. Mean loadings on the method factors were also positive. Mean loadings of the model applied to one-condition data were comparable to those applied to one-condition incentive data.

Convergent and discriminant validity.

Factor scores for the two-condition estimates of M and FP and one-condition estimates of FP entered into SPSS. Factor determinacy values for the two sets of factor scores were greater than .87 for all factors. The correlation between one- and two-condition estimates of FP was .667 for item indicators and .945 for parcel indicators indicating high convergent validity between the one- and two-condition factor scores, particularly when parcels were indicators. The correlation of the two-condition estimates of FP with difference scores were .574 and .621 for item and parcel indicators respectively. Analogous correlations for one-condition estimates were .367 and .525 for item and parcel indicators ($p < .001$ for all). These results indicated convergent validity with difference scores, again particularly for models based on parcels. Table 5 presents correlations of the factor scores with the BIDR SD and IM measures. With one exception, all of the correlations of M and FP with the BIDR measures were positive suggesting that there is moderate convergent validity between faking as measured here and the two measures of social desirability.

Surprisingly the correlation of M and FP was .665 and .637 for items and parcels respectively ($p < .001$) suggesting that the method variance in the honest response condition may also have contained a faking component.

Table 6 presents correlations of WPT scores with the various measures of faking. Although WPT correlated -.161 ($p < .05$) with SD, none of the other correlations of WPT scores with factor measures, difference scores or IM was significantly different from zero.

DISCUSSION

This study explored the convergent and discriminant validity of measures of faking based on common method factors within CFA models. This is the first study using an incentive-to-fake manipulation to examine the relationships between method factor, difference score, and social desirability measures of faking propensity. Positive correlations of the factor measures with difference scores and social desirability measures indicated convergent validity with these previously used measures suggesting that the factor measures, particularly those based on parcels, provide an alternative to the use of social desirability measures and difference scores. In addition to being valid measures of faking propensity, the one-condition factor scores are completely noninvasive to job applicants, thus promising to be a utilitarian measure of faking for use in operational settings.

The positive correlation of M and FP suggests that method bias in self-report questionnaires may include a large faking or impression management component. In this study, all participants experienced the honest response first, with no indication that an incentive would be given for the following condition. This suggests that even when there is no explicit instruction or incentive to fake, some respondents to self-report questionnaires engage in a modest amount of impression management to the extent that individual differences in the response distortion are measurable using the CFA techniques presented here.

This study also provided some evidence for discriminant validity of the measures. Although other studies have found faking ability to be positively related to cognitive ability (e.g., Biderman & Nguyen, 2004; Kuncel & Borneman, 2007), the results of this study suggest that faking propensity is not correlated with cognitive ability. This finding provides support for the proposition made in earlier applicant faking research that whereas faking ability was typically correlated with a measure of cognitive ability, faking propensity is correlated with measures of non-cognitive ability (Snell, A.F., Sydell, E.J., & Lueke, S.B., 1999).

It appears that there are only two criterion-related validity studies supporting the use of method factors similar to those presented here. In the Cellar et al. (1996) study, factor scores were found to have no significant correlation with supervisory ratings of training performance of flight attendant trainees. On the other hand, Biderman, Nguyen, Mullins, & Luna (2008) found that method factor scores were significantly related to supervisor ratings on three criteria for employees of a financial services company. This last result concurs with the suggestion of Morgeson, Campion, Dipboye, Murphy, & Schmitt (2007, p. 708) who noted that “whatever contributes to faking may also contribute to job performance – especially when one employs a supervisory rating as the criterion as is so often the case.”

One of the main drawbacks to the study of faking has been the lack of a useful measure of faking obtainable in applicant settings. As McFarland and Ryan (2006) noted, “It is critical that a construct-valid measure of faking behavior be developed for applicant samples.” Although measures of social desirability were often used in such contexts, the use of social desirability measures requires administration of an additional questionnaire and the inference that exaggeration on those questionnaires carries over to exaggeration on other personality dimensions in the applicant setting. Similarly, difference scores, while undeniably valid as measures of faking, are not feasible in most applicant settings. It would appear that the method factor measures considered here, long neglected in the faking literature, may be usable in such a context.

REFERENCES

Biderman, M. D., & Nguyen, N. T. (2004). *Structural equation models of faking ability in repeated measures designs.* Paper presented at the 19th Annual Society for Industrial and Organizational Psychology Conference, Chicago, IL.

Biderman, M. D., Nguyen, N. T., Mullins, B., & Luna, J. (2008, April). *A method factor predictor of performance ratings.* Paper presented at the 23rd annual conference of The Society for Industrial and Organizational Psychology, San Francisco, CA.

Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the five factor model. *Journal of Applied Psychology, 81*, 694-704.

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.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issue of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe,* Vol. 7 (pp. 1-28). Tilburg, The Netherlands: Tilburg University Press.

Griffith, R.L., Chmielowski, T., Snell, A.F., Frei, R.L., McDaniel, M.A. (2000) Does Faking Matter? An Examination of Rank Order Changes in Applicant Data. Symposium presented at the 15th annual meeting of the Society of Industrial and Organizational Psychologists. New Orleans, LA.

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

Jo, M., Nelson, J. E., & Kiecker, P. (1997). A model for controlling social desirability bias by direct and indirect questioning. *Marketing Letters, 8*, 429-437.

Jo, M. (2000). Controlling social desirability bias via method factors of direct and indirect questionining in structural equation models. *Psychology & Marketing, 17*, 137-148.

Kenny, D. A., & McCoach, D. B. (2003). Effect of number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333-351.

Klehe, U., Kleinmann, M., Hartstein, T., Melchers, K. G., & König, C. J. (2008). Fabulous fakers on the Big Five: The role of candidates' ability to identify criteria in responding to personality assessments. Manuscript submitted for publication.

Kuncel, N. R., & Borneman, M. J. (2007) Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment, 15*, 220-231.

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

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*, 979-1016.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., III. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88,* 348-355.

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus User's Guide. 5th Ed. Los Angeles, CA: Muthén & Muthén.

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

Paulhus, D.L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, and L.S. Wrightsman (Eds.). *Measurement of Personality and Social Psychological attitudes.* San Diego: Academic Press, pp. 17-59.

Podsakoff, P.M., MacKenzie, S. B., Lee, J., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903.

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

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.

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

Spector, P. E. (1992). *Summated rating Scale Construction: An Introduction.* Newbury Park, CA: Sage.

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

Wonderlic, Inc. (1999). *Wonderlic's Personnel Test manual and scoring guide*. Chicago: IL: Author.

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, *20*, 71-87.

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

Footnotes

[1] As shown in Figure 2, following Biderman and Nguyen (2004) identical indicators from the honest and incentive conditions were allowed to correlate. In addition, incentive condition residuals within dimensions were allowed to correlate. Allowing these correlated residuals improved goodness-of-fit but did not change conclusions concerning relative fit of models or correlations involving factor scores.

Table 1.

Means, standard deviations, and correlations of whole-scale scores for the Big Five dimensions.

H scales are from the honest condition. F scales are from the incentive condition.

Dimension	1	2	3	4	5	6	7	8	9	10	11	12
1. H E	(.845)											
2. H A	.275	(.788)										
3. H C	.035	.347	(.778)									
4. H S	.207	.163	.057	(.861)								
5. H O	.159	.276	.165	.226	(.766)							
6. F E	.853	.253	.029	.164	.150	(.878)						
7. F A	.211	.691	.189	.091	.169	.342	(.810)					
8. F C	.080	.261	.690	.095	.120	.212	.411	(.824)				
9. F S	.208	.117	.046	.798	.181	.258	.229	.288	(.883)			
10. F O	.189	.295	.035	.143	.751	.328	.392	.302	.261	(.774)		
11. SD	.282	.153	.279	.549	.284	.266	.114	.243	.458	.182	(.668)	
12. IM	-.104	.189	.300	.283	.011	-.062	.187	.291	.220	.000	.224	(.781)
Mean	4.507	5.250	4.793	3.979	4.729	4.622	5.354	5.025	4.440	4.825	4.261	3.579
SD	0.974	0.822	0.805	1.095	0.730	1.011	0.812	0.853	1.133	0.750	0.591	0.789

Table 2

Goodness-of-fit statistics of general and special models applied to two-condition data..

Individual Items as Indicators

Statistic	Model: M, FP Estimated	FP omitted	M Omitted	M, FP Omitted
χ^2	7896.791 [a]	8252.524 [a]	8155.427 [a]	8596.757[a]
CFI	.737	.713	.721	.672
RMSEA	.062	.064	.063	.064
$\Delta\chi^2$		355.733[a]	258.636 [a]	699.966[a]
Model df	4464	4515	4515	4565
Δ df		51	51	101
Two-item parcels as indicators				
χ^2	1743.461[a]	1979.114[a]	1854.274[a]	2201.553
CFI	.897	.866	.885	.837
RMSEA	.058	.065	.061	.071
$\Delta\chi^2$		235.653 [a]	110.813[a]	458.092 [a]
Model df	1039	1065	1065	1090
Δ df		26	26	51

[a] $p < .001$

Table 3.

Goodness-of-fit statistics of general and special models applied to one-condition data

.	Model	
Statistic	FP Estimated	FP omitted
Individual items as indicators		
χ^2	2190.702 [a]	2486.204 [a]
CFI	.758	.703
RMSEA	.069	.075
$\Delta\chi^2$		355.733 [a]
Model df	1115	1165
Δ df		50
Two-item parcels as indicators		
χ^2	466.311 [a]	594.442 [a]
CFI	.907	.864
RMSEA	.068	.078
$\Delta\chi^2$		128.131 [a]
Model df	240	265
Δ df	25	

[a] $p < .001$

Table 4.

Mean loading estimates of latent variable indicators from Model 1.

	Two-condition model		One-condition model	
Latent Variable	Items	Parcels	Items	Parcels
Extraversion (E)				
Honest Condition	.56	.69		
Incentive Condition	.54	.67	.56	.65
Agreeableness (A)				
Honest Condition	.11	.43		
Incentive Condition	.11	.39	.51	.54
Conscientiousness (C)				
Honest Condition	.44	.53		
Incentive Condition	.37	.41	.54	.48
Stability (S)				
Honest Condition	.59	.70		
Incentive Condition	54	.59	.61	.56
Openness (O)				
Honest Condition	.40	.55		
Incentive Condition	.33	.45	.14	.52
M Honest Condition	.29	.34		
FP Incentive Condition	.37	.43	.29	.45

Table 5

Correlations of factor measures of faking propensity with social desirability measures.

	Two-condition				One-condition	
	Items		Parcels		Items	Parcels
	M	FP	M	FP	FP	FP
SD	.295[c]	.225[b]	.447[c]	.308[c]	.255[c]	.438[c]
IM	.197[b]	.175[a]	.302[c]	.243[b]	-.017	.247[c]

[a] $p < .05$ [b] $p < .01$ [c] $p < .001$

Table 6

Correlations of factor and social desirability measures with the WPT.

	Two-condition				One-condition		BIDR	
	Items		Parcels		Items	Parcels		
	M	FP	M	FP	FP	FP	SD	IM
WPT	.030	.063	.010	.071	.048	.032	-.161[a]	-.038

[a] $p < .05$

Figure Captions

Figure 1. A method/faking factor added to a confirmatory factor analysis of Big Five questionnaire items.

Figure 2. A method/faking factor model applied to two-condition data. The model was applied to five two-item parcels per dimension, as shown, and also to 10 individual item indicators per dimension (not shown). To improve clarity, residuals are not shown. Indicators beginning with H represent the honest condition, and those beginning with I represent the incentive condition. As illustrated for extraversion, pairwise correlations between corresponding honest and faking indicators were estimated for all Big Five dimensions. As illustrated for conscientiousness, faking condition residuals within each Big Five dimension were allowed to covary.

Figure 1

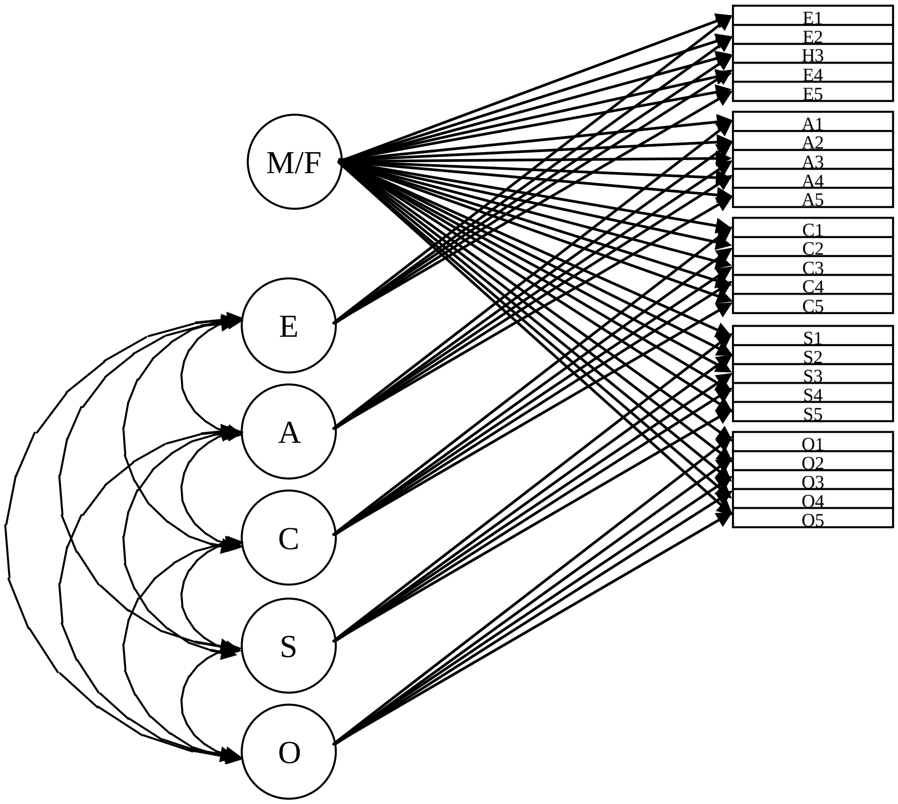
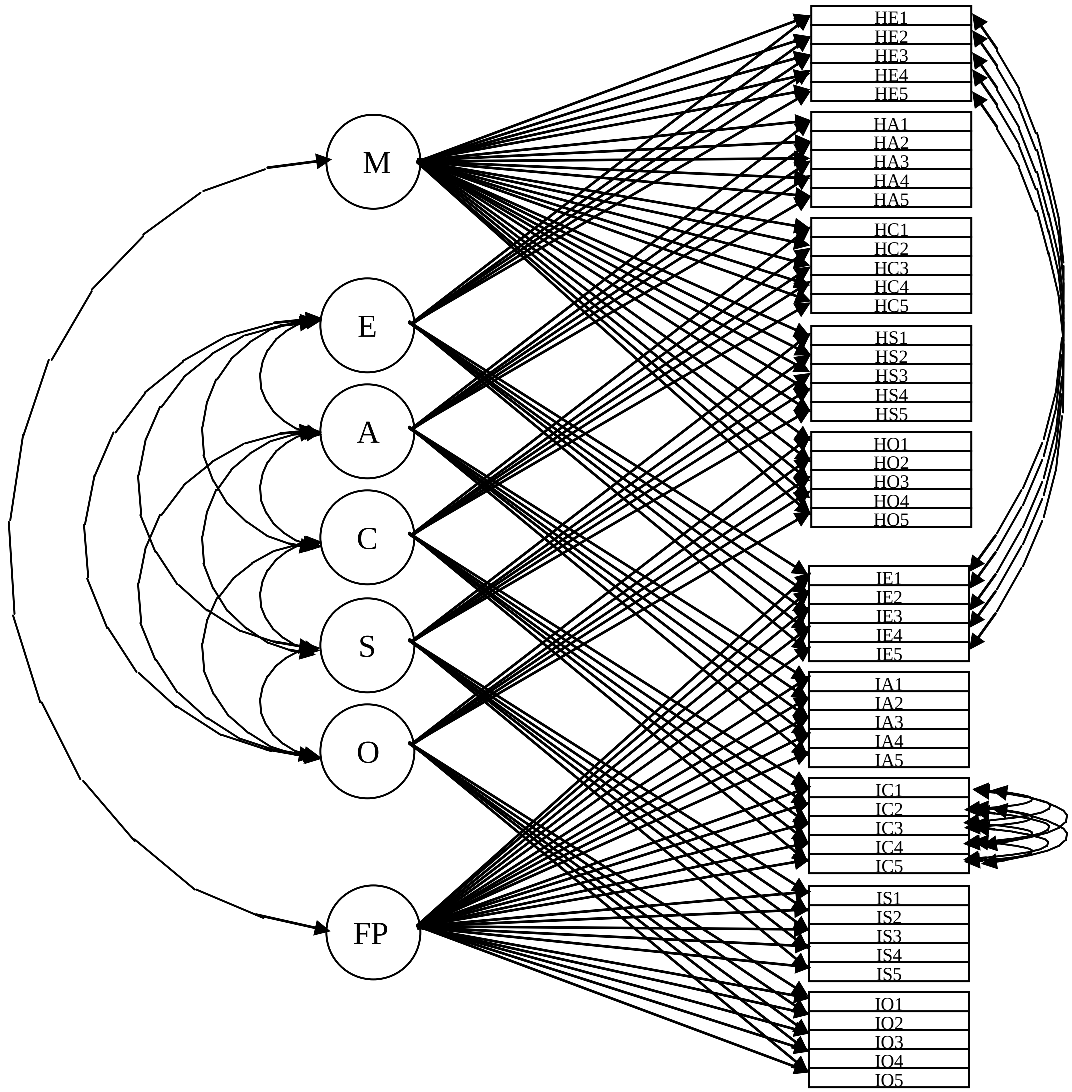

M/F
E
A
C
S
O
E1
E2
H3
E4
E5
A1
A2
A3
A4
A5
C1
C2
C3
C4
C5
S1
S2
S3
S4
S5
O1
O2
O3
O4
O5

Figure 2


M
E
A
C
S
O
FP
HE1
HE2
HE3
HE4
HE5
HA1
HA2
HA3
HA4
HA5
HC1
HC2
HC3
HC4
HC5
HS1
HS2
HS3
HS4
HS5
HO1
HO2
HO3
HO4
HO5
IE1
IE2
IE3
IE4
IE5
IA1
IA2
IA3
IA4
IA5
IC1
IC2
IC3
IC4
IC5
IS1
IS2
IS3
IS4
IS5
IO1
IO2
IO3
IO4
IO5