Structural equation models of faking ability in repeated measures designs

Michael D. Biderman

University of Tennessee at Chattanooga
Department of Psychology
615 McCallie Ave.,
Chattanooga, TN 37403
Tel.: (423) 755-4268
Fax: (423) 267-2289
E-mail: Michael-Biderman@utc.edu

Nhung T. Nguyen

Lamar University
Department of Management & Marketing
4400 Martin Luther King Parkway
Beaumont, TX 77706
Tel.: (409) 880-8295
Fax: (409) 880-8620
E-mail: nguyennt@hal.lamar.edu

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. E-mail: Michael-Biderman@utc.edu

Paper presented at the 19th Annual Society for Industrial and Organizational Psychology Conference, Chicago, IL, 2004.

Poster

TITLE

Structural equation models of faking ability in repeated measures designs

ABSTRACT

Models were compared on data in which a situational judgment test and measures of the Big Five were administered under honest and fake good instructions. A model with latent variables representing the six measures and a latent variable representing faking ability proved to be a useful representation of the data.

PRESS PARAGRAPH

Many experimental studies of applicant faking have employed the difference between test performance when instructed to fake with performance when instructed to respond honestly as a measure of faking ability. There have been many arguments against using such difference scores as measures of change in the psychological literature. This study investigated an alternative approach to the conceptualization of behavior in such studies using structural equation modeling techniques. The model developed provided a reasonable fit to the data and allowed an examination of applicant faking that could not have been carried out as efficiently using the difference score approach.

Structural equation models of faking ability in repeated measures designs

One result of the increase in importance of noncognitive measures in selection processes has been a renewed focus on applicant faking and its implications for selection processes. The issue of applicant faking is of little importance when cognitive tests are used for selection. However, when noncognitive tests, such as personality tests, are used, the possibility of choosing employees because of their ability to fake appropriate scores on such tests rather than their actual personality characteristics presents a dilemma for selection specialists. Of course, if the ability to fake is a predictor of performance or other attributes important for the job, the issue of faking is less problematic. And if the amount of faking is constant among the applicant population, then, again, the existence of such faking will not present a problem in personnel selection. But if there are individual differences in faking across the applicant population and if that faking is not related to important job attributes, then the potential impact of such faking on validity of selection instruments is of concern to the selection specialist.

The measurement of faking behavior depends on the experimental design in which it is studied. In studies in which no manipulation of faking instruction is employed, investigators have typically used separate measures of faking, such as social desirability scales (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Rosse, Stecher, Miller, & Levine, 1998). In other studies, participants have been instructed to respond honestly at one point in the research and to respond in a faking good fashion in another part. In these repeated measures designs, difference scores have been used to measure the extent of faking (e.g., McFarland & Ryan, 2000, 2003).

In addition to differences in the operationalization of faking behavior, some studies have focused on the tendency of respondents to fake in various contexts (e.g., applicant vs. job incumbent; Hough et al., 1990), while other, recent studies have examined the ability to fake regardless of context (e.g., McFarland & Ryan, 2000; 2003). The focus of the present research is the ability of applicants to fake. Following McFarland and Ryan (2000, 2003) we employed a repeated measures design. Specifically, we investigated structural equation models as alternatives to the use of difference scores as measures of faking in such designs.

The difference between test performance obtained under conditions to fake good and the performance under instructions to respond honestly, referred to here as F-H, is an example of a difference score encountered in a variety of research contexts within psychology. Difference scores have been employed as independent variables in studies to represent person-environment fit, value fulfillment, met expectations, and self-other agreement to name just a few (Edwards, 2002). They also have been considered as dependent variables in studies investigating faking-personality relationships (e.g., McFarland & Ryan, 2000). As dependent variables, difference score measures of faking are in good company with measures of pre-post change in the two-level repeated measures designs (e.g., Cribbie, & Jamieson, 2000). The models investigated in this study are not restrictive concerning the independent/dependent variable role of applicant faking.

Many arguments have been presented against the use of difference scores, i.e., F-H, as both independent (Edwards, 2002; Cribbie & Jamieson, 2000) and dependent variables (Edwards, 1995) in psychological research. First of all, the predominant argument is the fact that in some circumstances they may have very low reliability (e.g., Edwards, 2002). Secondly, when treated as dependent variables (i.e., change scores) in pretest-posttest designs, it has been pointed out that such use is a special case of analysis of covariance in which within-groups slopes relating posttest to pretest are assumed to equal one (Pedhazur & Schmelkin, 1991). Thirdly, difference or change scores suffer from the fact that they are inherently inversely related to pretest values preventing examination of the precursors of change in some instances and sometimes leading to incorrect conclusions concerning such precursors (Cohen & Cohen, 1983).

In spite of the arguments against the use of difference scores, they have been employed in recent studies of applicant faking of noncognitive measures. For example, McFarland and Ryan (2000) found generally positive correlations among difference scores from NEO-PI measures of the Big Five personality dimensions (Costa and McCrea, 1989), a bio-data instrument, and an integrity test, leading to the suggestion that there are consistent individual differences in the extent of faking that are not test specific. Their study provided "initial evidence that even when people fake responses, there are individual differences in the extent to which people fake." (McFarlend & Ryan, 2000, p. 817-818). In spite of these positive results, there were certain tests that could not be conducted as efficiently as might have been desired using difference scores. These were tests of the relationship of personality traits to faking. Since faking was defined as the F-H test performance difference faking scores for each personality trait were negatively correlated with the corresponding trait scores preventing a clear cut examination of the relationship of faking ability to respondent attributes as represented by the Big Five personality traits. Moreover, even though McFarland and Ryan (2000) found evidence consistent with the existence of a single individual difference variable representing faking ability, their analyses were conducted separately using faking scores for each personality dimension. An approach, which allows examination of a single faking ability variable, would have been desirable.

The present study was designed to address the issue of difference scores as a measure of applicant faking and whether there existed a single faking ability trait through the use of structural equation modeling techniques. The data to which the model was applied (Nguyen, 2002) consisted of noncognitive measures administered twice, once under instructions to respond honestly and again under instructions to fake good. A series of structural equation models was applied to the data. Models incorporating the relationships of a faking latent variable to latent variables representing personality dimensions and to cognitive ability were assessed.

Method

Two hundred three undergraduate and graduate students from two southeastern public universities participated in the study in exchange for partial course credit. All participants were given the Wonderlic Personnel test (Form A). After that all participants completed a 31 item situational judgment test (SJT) and a 50-item instrument measuring the Big Five personality dimensions twice, once under instructions to respond honestly and also under instructions to respond faking good. The order of instructions was counterbalanced with half of the participants completing the measures under the “honest” instructions first and the other half under the “faking good” instructions first. The SJT was the Work Judgment Survey described by Smith and McDaniel (1998). The test consists of 31 problem scenarios or situations with five possible responses. Empirical keying had identified one response as the “Most likely” response to the situation and another response as the “Least likely” response. Respondents were asked to indicate which response they would most likely make and which response they would least likely make to the situation. Estimates of internal consistency reliability were .74 and .78 for the Honest and Faking instructional conditions respectively. We used the Big 5 personality inventory developed by Goldberg (http://ipip.ori.org/ipip). Each dimension was indicated by 10 Likert-type items. Respondents were asked to indicate the accuracy of each item as a descriptor of themselves with a number ranging from 1 (*very inaccurate*) to 5 (*very accurate*). The scales have been validated against other established scales (e.g., NEO-PI) and shown to have good reliabilities (Goldberg, 1999; in press). Estimates of reliability ranged from .74 to .90.

In order to provide adequate indicators of each of the latent variables in the model, three "testlets" for the SJT and for each of the personality dimensions were formed. Bandalos (2002) found that the use of item testlets or parcels resulted in better fitting solutions when items had a unidimensional structure. Previous work with the scales used here suggested that each was essentially unidimensional, supporting the use of such testlets. To increase the likelihood that the testlets for each measure would have equal variances, all the testlets for each measure had the same number of items. For the SJT the last item in the scale was dropped and three 10-item testlets for each instructional condition were formed. For the personality dimensions, the last item in each scale was dropped and three three-item testlets were formed for each scale under each instructional condition. This resulted in 36 testlets, with three from each measure under the Honest instructional condition and three from each measure under the Faking instructional condition. (The correlation matrix of the testlets is available from the first author on request.) The models were applied to the data using Amos, Version 4. (Arbuckle & Wothke, 1999).

Results

The model applied to the data is similar in conceptualization to models of pretest-posttest data presented by Cribbie and Jamieson (2000, p. 902). The Cribbie and Jamieson model contained two latent variables. One represented pretest ability, with both observed pretest scores and posttest scores as indicators. Posttest scores were included as indicators of the pretest latent variable since presumably the observed posttest scores were influenced by some of the factors affecting pretest performance. The second latent variable, presumably influenced only by factors occurring between pre- and posttest , was indicated only by posttest scores. The model presented here is analogous to Cribbie and Jamieson's (2000) model with the noncognitive dimensions serving the same role as Cribbie and Jamieson's pretest performance and faking serving in the same role as Cribbie and Jamieson's posttest performance. It is different from Cribbie and Jamieson's (2000) model in that the there were multiple latent variables in the role analogous to pretest performance.

The procedure followed involved first creating a measurement model involving the two types of latent variables – noncognitive dimensions and faking – then testing structural relationships using that measurement model. The first model, presented in Figure 1, was a simple confirmatory factor analytic model with six latent variables representing the six noncognitive dimension – the SJT and the Big Five dimensions. Each latent variable was indicated by six testlets – three obtained under Honest instructions and three under Faking instructions. Table 1 presents four popular goodness-of-fit statistics for the several models investigated here – the chi-square statistic, the GFI, the AGFI, and the RMSEA (Fan, Thompson, & Wang, 1999). Figure 1 is taken from the Amos output and shows variable names used in the original dataset. Because of the complexity of the path diagram in Figure 1, a schematized representation of the path diagram of models presented later has been employed. In this schematized diagram, one symbol is used for each set of three testlets, and only one path from each latent variable to the symbol representing a set of indicators is represented. Values of loadings on each three-testlet triplet are represented by the mean of the three loadings of the testlets within the triplet.

Inspection of Figure 1 shows that all testlets loaded quite highly on their respective latent variables. The goodness-of-fit values presented in Table 1, however, suggest that this model is not a good fit to the data. Model 1 does not contain any parameters to account for faking. Since half the indicators were measured under instructions to fake, the possibility that individual differences in faking could account for some of the lack of fit of the "No faking" model was investigated next. Faking ability was added to the model as a single latent variable indicated by the testlets from only the Faking instructional set. This model is presented in Figure 2. Since Model 2 is a generalization of Model 1, a chi-square difference test can be used to test the hypothesis that all regression weights from the Faking latent variable to its indicators are zero. The chi-square difference value 918.33 with df=18, $p<.001$, indicating that addition of the Faking latent variable resulted in a significant improvement in model fit. The other goodness-of-fit statistics reported in Table 1 support this conclusion.

Although the improvement in fit of Model 2 relative to Model 1 suggested that the addition of a Faking Ability latent variable was an appropriate modification to the model, the goodness-of-fit statistics in Table 1 were not in the range considered acceptable for such statistics. Given the preliminary nature of this investigation, we felt that some guidance from the data would be appropriate to suggest areas of improvement in the model. Inspection of the modification indices for Model 2 revealed many positive modification index values for covariances between the error terms within triplets of testlets from the Faking condition. Based on this post hoc analysis of the modification indices, covariances between error terms for the Faking condition testlets were added to the model, forming Model 3. The model is presented in Figure 3. As can be seen in Table 1, the resulting decrease in the chi-square statistic was significant ($X^2(18)= 266.41$, $p< .001$).

Since the testlets were in pairs, with each Honest condition testlet consisting of the same items as a Faking condition testlet, we expected that there would be correlations between pairs of testlets from each condition that might not be represented by the loadings on the latent variables. For example, we expected higher correlations between Honest condition Testlet 1 scores and Faking condition Testlet 1 scores than could be represented by the loadings on the latent variables indicated by the testlets and the faking latent variable. Inspection of the modifation indices supported this belief - in general, the modification indices for the correlation between Honest and Faking condition testlets consisting of the same items were large and positive. For this reason, a fourth measurement model was formed, allowing covariances between the three corresponding pairs of testlets for the SJT latent variable and for the Big 5 measures. The fit of this model, presented in Figure 4, was substantially better than the fit of Model 3 ($X^2(18) = 237.36$, $p < .001$). Moreover the goodness-of-fit statistics were close to recommended values for these statistics (e.g., Kelloway, 1998, p. 27-28).

The fit of Model 4 is marginally acceptable, although at this stage in the investigation of this model it is promising. Inspection of the estimates revealed that all loadings of indicators from both the Honest and Fake conditions onto the SJT and Big Five latent variables were positive, as one would expect. The loadings of indicators from the Faking conditions onto the Faking ability latent variable were all positive, again in line with our expectations. The fact that all the loadings were positive supports the conclusion of McFarland and Ryan (2000) that there are consistent individual differences in faking ability.

Model 4 formed the basis for the following tests of hypotheses concerning relationships of faking to other variables . The model presented here is both a model for both the measures upon which faking is based, i.e., the SJT and Big 5 measures, and a model for faking of those measures. For that reason, it may be used to investigate both the interrelationships of the dimensions and the relationship of faking to those dimensions, although our interest here is on the latter. And, unlike a difference score analysis, in which a negative correlation between the measure of faking and the measures upon which faking is based is built into the faking measure and therefore cannot be tested, the nature of the relationships between the faking latent variable in this model and the SJT and Big Five measures is testable. Moreover, the relationships of the faking latent variable to other variables outside the model, such as cognitive ability, can also be tested within the framework of the model presented here.

To show the versatility of this model, three sets of relationships were tested through the addition of structural paths to the model. In the first, the extent to which Faking Ability was related to the several personality dimensions was investigated by adding to the model regression weights of the faking latent variable onto the Big Five measures. This addition was suggested by the analyses of McFarland and Ryan (2000) in which the relationship of faking based on difference scores to selected personality dimensions was investigated. (We saw no theoretical reason for supposing that faking ability would be related to SJT performance and didn't include a regression to the SJT latent variable.) Secondly, on the assumption that faking ability, like other abilities, would be related to cognitive ability, we tested the relationship of faking ability to cognitive ability by adding a regression weight of the faking latent variable onto the Wonderlic scores. Finally, in order to investigate the possibility of an order effect in faking, a variable representing the order of the instructional conditions was included and faking ability regressed on to it. The resulting structural equation model, Model 5, is presented in Figure 5.

Since Model 5 is not a simple generalization of the previous model, no overall comparison of goodness-of-fit with the previous models was conducted. In fact the chi-square goodness of fit statistic for this model is larger than that for Model 4 because of the addition of the observed cognitive ability and order variables to data to be fit by the model. We note however, that the other indices of fit were about equal to those of Model 4, suggesting that the structural model is not fitting the data noticeably worse than the measurement Model 4. The relationships of faking ability to the variables within the measurement model (the Big 5 latent variables) and to the variables added to the model were assessed by interpreting the individual regression coefficients added to the model. These coefficients are in italics in Figure 5 to distinguish them from the other values that represent means of three coefficients. None of the regression weights associating the faking ability latent variable with the Big 5 latent variables was significantly different from 0. That is, when controlling for cognitive ability and order, faking ability is not related to any of the Big 5 personality dimensions. However, faking ability was positively related to cognitive ability ($p = .011$). Those higher in cognitive ability showed greater faking ability. Finally, the weight relating faking to order of presentation was not significantly different from zero, suggesting that faking ability was not related to order of instructions to respond honestly or to fake good.

Discussion

Although certain characteristics of the measurement model developed here, specifically the need for correlated error terms, require further study, the goodness-of-fit statistics were near those values that represent good fit to the data (Kelloway, 1998). Assuming measurement model characteristics can be accounted for, the model represents a promising way of investigating simultaneously the relationships among noncognitive dimensions and the ability to fake on those dimensions. It also shows how the relationships of other variables to faking can be efficiently studied.

The model presented here differs from analyses based on difference scores in that faking ability, as represented by a latent variable in the model, is directly indicated only by performance in the Faking condition. On the contrary, a difference score conceptualization would require that the faking latent variable be indicated by performance under both Honest and Faking conditions. Moreover, a strict difference score model would require regression weights equal in absolute value but with opposite signs to the Honest condition and Faking condition indicators, respectively. Such a conceptualization was beyond the scope of the present investigation, but a complete examination of the model presented here will certainly require development of models utilizing opposite-weight indicators to provide comparisons with the conceptualization presented here.

In this model, only one faking ability latent variable was included. Fit could have been improved by the inclusion of separate faking ability latent variables for each dimension measured. Of course, the existence of only one faking ability is a much more parsimonious possibility than the existence of separate abilities associated with each personality domain. In fact, differences in loadings of the faking ability latent variable onto the testlets indicating difference dimensions might suffice to account for differences in fakability of different measures. The differences in loadings of the testlets onto the faking latent variable suggest that the single faking ability measured here is differentially manifested across the different variables. It suggests, for example, that the SJT test was the least fakable of the six measures and that Conscientiousness, Emotional Stability, and Openness were the most susceptible to faking.

The finding that smarter people are better fakers confirms previous researchers' proposal (e.g., Snell, Sydell, & Lueke, 1999) that cognitive ability predicts faking. However, our study is the first to empirically test this relationship. This finding represents an interesting dilemma for selection specialists. On the one hand, selection measures uncontaminated by faking are probably most desirable. For that reason, suspicion has been cast on the use of noncognitive tests such as the tests employed here because of their potential contamination by faking. But, since cognitive ability is the best single predictor of performance on a variety of jobs (Schmidt & Hunter, 1998) the fact that scores on fakable tests might be contaminated by an ability related to cognitive ability suggests that using use of such tests might not have the deleterious effects on validity that have been feared. Certainly, a means of isolating both the determinants of faking ability and the effects of faking ability in selection contexts is one step toward better understanding of this dilemma. We hope that the model presented here represents such a step.

References

Arbuckle, J. A., & Wothke, W. (1999). *Amos 4.0 User's Guide.* Chicago: The SmallWaters Corporation.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9(1)*, 78-102.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Costa, P.T., & McCrae, R.R. (1989). *The NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Cribbie, R.A., & Jamieson, J. (2000). Structural equation models and the regression bias for measuring correlates of change. *Educational and Psychological Measurement, 60,* 893-907.

Edwards, J.R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64*, 307-324.

Edwards, J.R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In Drasgow, F., & Schmitt, N. (Eds.). *Measuring and Analyzing behavior in organizations: Advances in measurement and data analysis*. Jossey-Bass.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6(1)*, 56-83.

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. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R. (in press). The comparative validity of adult personality inventories: Applications of a consumer-testing framework. In S. R. Briggs, J. M. Cheek, & E. M. Donahue (Eds.), *Handbook of Adult Personality Inventories.*

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.

Http://ipip.ori.org/ipip. *International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences.*

Kelloway, E. K. (1998). *Using LISREL for Structural Equation Modeling: A Researcher's Guide.* Thousand Oaks, CA: Sage.

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

McFarland, L. A., & Ryan, A. M. (2003). Toward an integrated model of applicant faking behavior. Manuscript submitted for publication.

Nguyen, N.T. (2002). *Faking in situational judgment tests: An empirical investigation of the work judgment survey.* Dissertation Abstracts International Section A: Humanities & Social Sciences, Vol 62(9-A), pp. 3109.

Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, Design, and Analysis: An Integrated Approach.* Hillsdale, NJ: Lawrence Erlbaum.

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.

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

Smith, K. C., & McDaniel, M. A. (April, 1998). *Criterion and construct validity evidence for a situational judgment measure.* Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

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

Table 1 Goodness of fit statistics.

Model	Degrees of freedom	Chi-square	Chi-square Difference	P <	GFI	AGFI	RMSEA
1	579	2241.38			.558	.492	.119
2	561	1323.05	918.33	.001	.736	.687	.082
3	543	1056.64	266.41	.001	.778	.728	.068
4	525	819.28	237.36	.001	.823	.776	.053
5	585	900.23			.818	.770	.052

Each of models 2 through 4 was a generalization of the previous model.

Figure 1. Model 1: Confirmatory Factor Analysis Model. Six latent variables representing the SJT test and the Big Five measures with six indicators each, three from Honest instruction testlets and three from Faking instruction testlets. Labels are: E=Extroversion; A=Agreeableness; C=Conscienciousness; S=(Emotional) Stability; O=Openness. Figure is from Amos output.

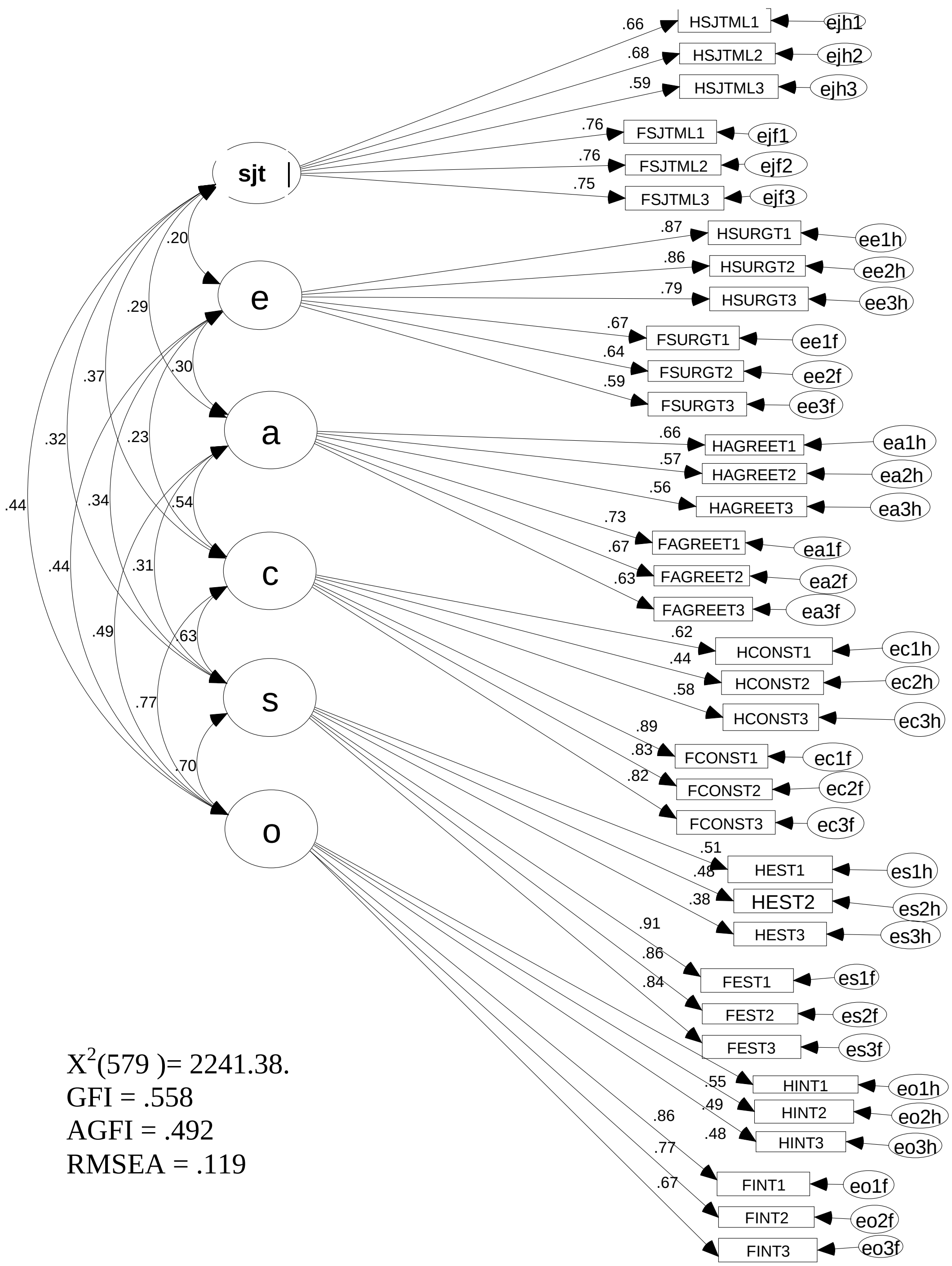
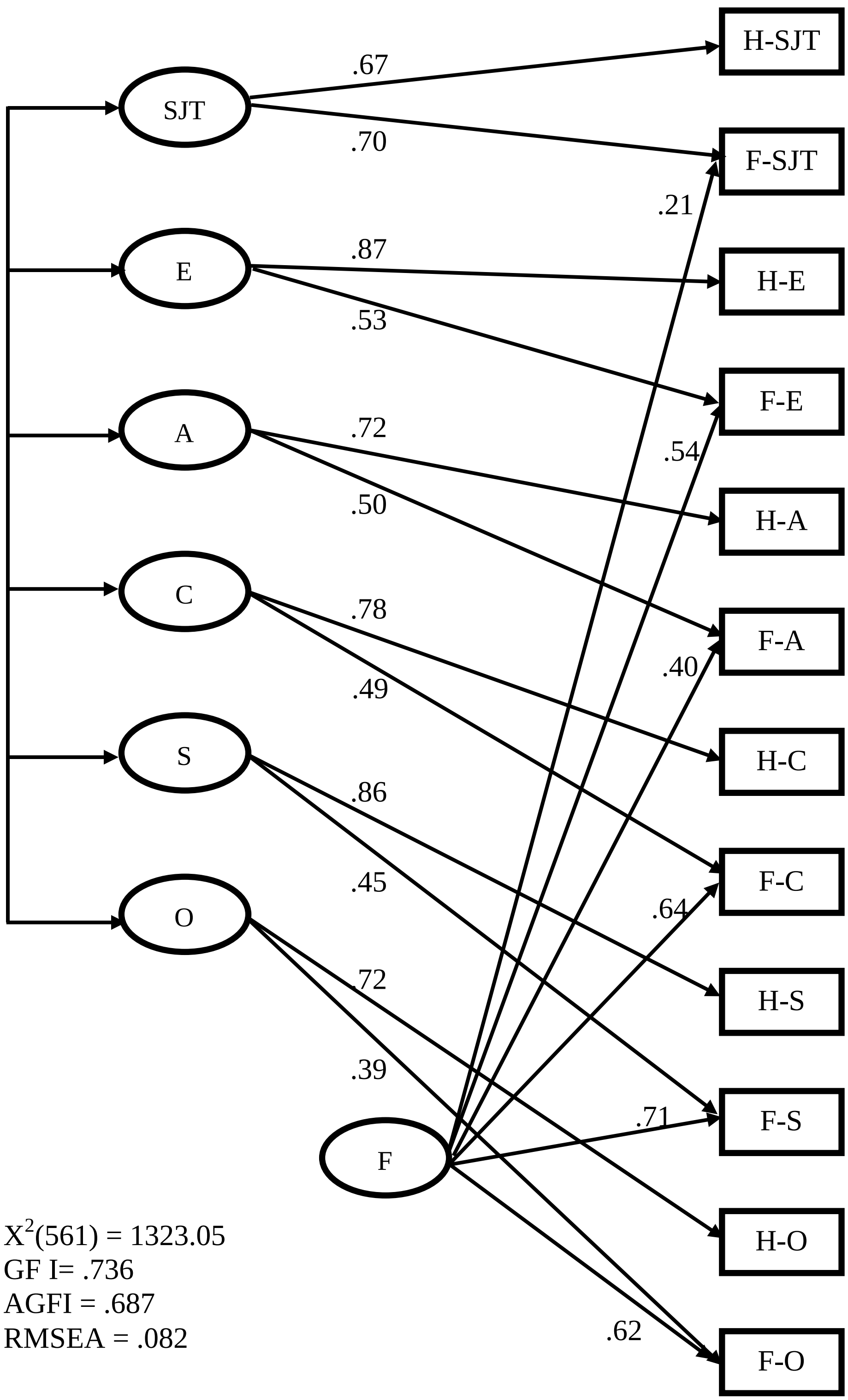


Figure 2. Model 2: Faking Model. Latent Variables represent the Six Measures + Faking Latent Variable. Each rectangle represents three testlets. Each loading is the mean of loadings of three testlets.



$X^2(561) = 1323.05$
GF I= .736
AGFI = .687
RMSEA = .082

Figure 3. Model 3: Faking Model with Correlated Faking Testlet Errors. Covariances between Faking Instruction Testlets errors were estimated. Values are means of loadings on three testlets or means of correlations between error terms.

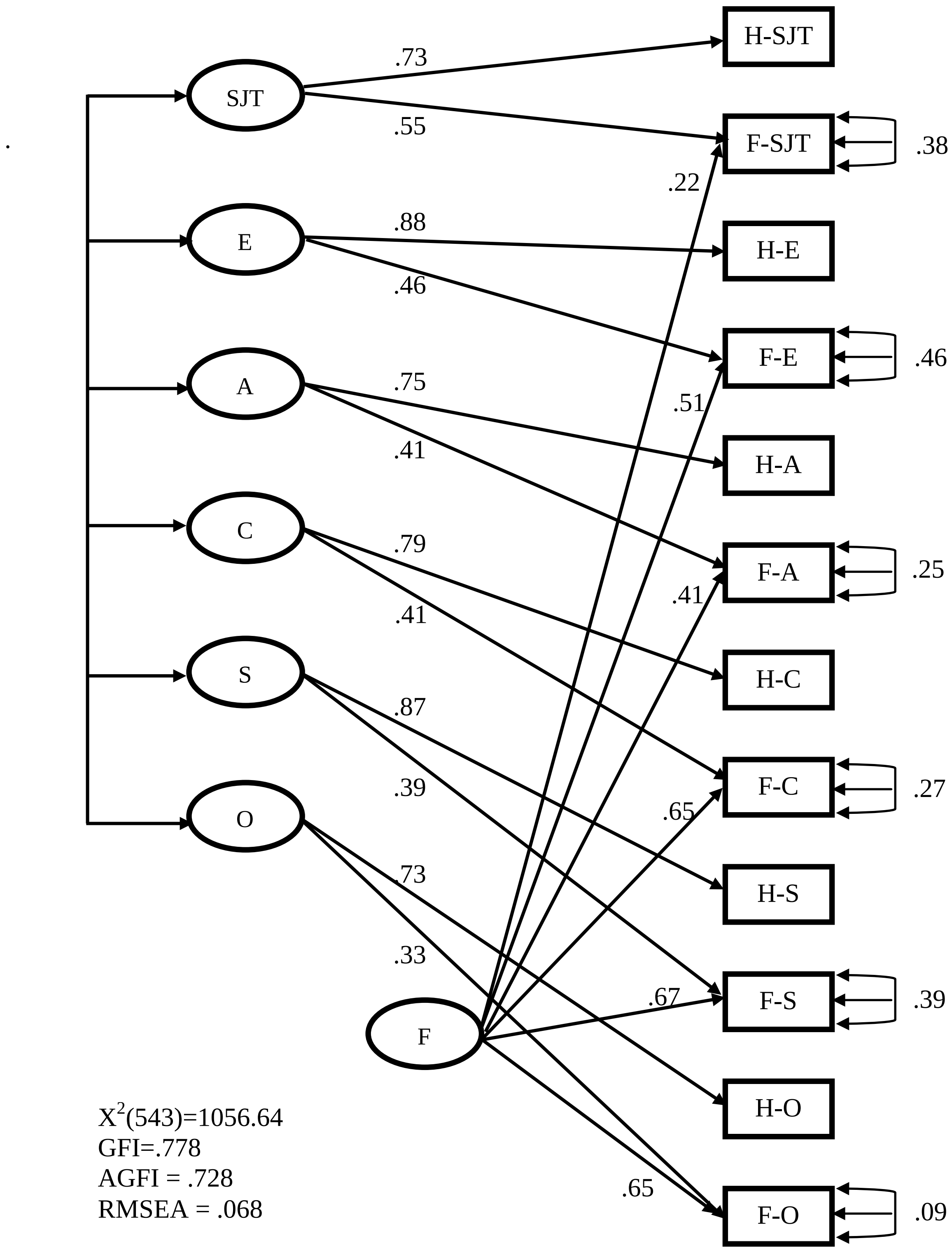


Figure 4: Model 4: Faking Model with Correlated Faking and Faking-to-Honest Testlet Errors. Covariances between Faking instruction testlets and between corresponding testlets from Honest and Faking Instructions were estimated. Values are means of loadings or means of correlations between testlet triplets.

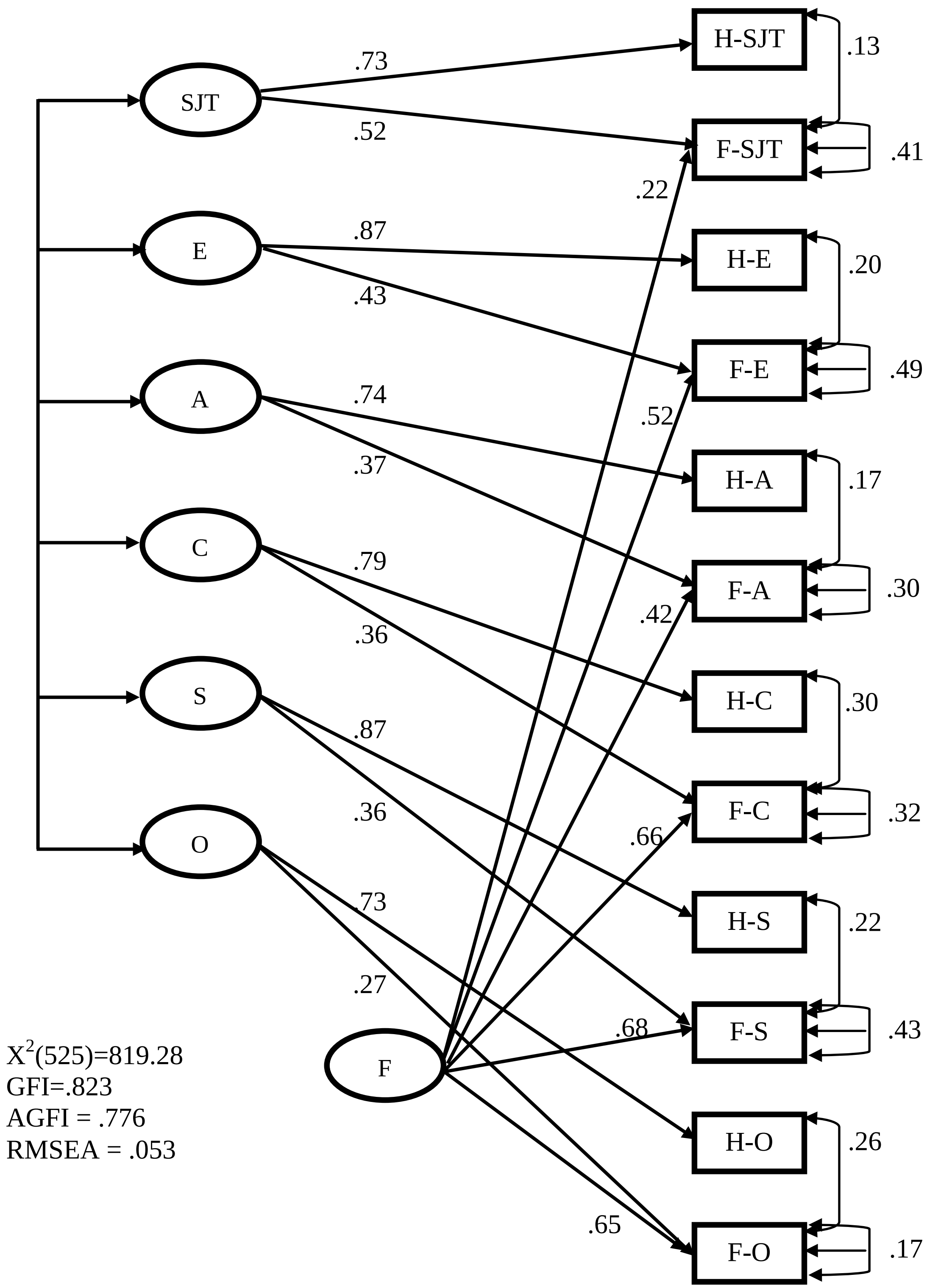


Figure 5. Model 5: Complete Faking Model with Tests of Structural Hypotheses. Model 4 measurement model with regression weights from the Big Five latent variables, the Wonderlic, and Order of receipt of Honest and Faking Instructions. Italicized values are individual standardized regression weights. (* represents p < .05.)