

Measuring response distortion using structural equation models.

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

Nhung T. Nguyen

Towson University
Department of Management
8000 York Road
Towson, MD 21252
Tel.: (410) 704-3417
Fax: (410) 704-3236

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 conference, New Directions in Psychological Measurement with Model-Based Approaches. February 17, 2006. Georgia Institute of Technology, Atlanta, GA.

The authors would like to thank Lyndsay B. Wrensen and J. Michael Clark for their assistance gathering the data for two of the studies reviewed here.

Measuring response distortion using structural equation models

The resurgence of personality tests in employee selection has generated a renewed interest in the measurement of applicant response distortion, or faking. Although both “faking good” and “faking bad” could be possible in personality testing, “faking good” has received more attention of organizational researchers. This emphasis stems from the linkage between applicants’ attempt to present them in a favorable light and increased likelihood of being hired (Nguyen & McDaniel, 2001). Many studies have documented the fakability of personality tests as well as the prevalence of faking in applicant samples. However, there has been less research on modeling applicant faking as a substantive construct (e.g., faking ability, faking propensity or motivation) and its relationship to the measurement properties of personality tests. Recently, Biderman and Nguyen (2004) proposed modeling faking ability via a structural equations model that shows promise for understanding and recognizing applicant faking. This paper reviews the applications of that model, explores the relationship of faking to method variance within the context of the model, and presents a way of using the model to measure response distortion from groups such as applicant populations.

Surrogate variable approaches. Early research on response distortion was characterized by the use of instruments whose main purpose was to measure the tendency to distort self-report to present a favorable image. Primary among these were measures of social desirability (e.g., Paulhus, 1984; 1991; Vasilopoulos, Reilly, & Leaman, 2000). The number of studies using social desirability scales was such that even now it is not uncommon to see distortions of responses that might be characterized as “faking good” described as socially desirable responding. Although the use of social desirability as a relatively pure indicator of faking had much face validity, this line of research met with limited success partially due to the difficulty in separating variance due to faking from variance due to self-deception as a personality trait (e.g., Ellingson, Sackett, & Hough, 1999).

An additional criticism of surrogate measures of response distortion is that they are necessarily indirect measures. When a personality inventory is used in an applicant population, the interest of the selection specialist is on distortion of the responses to that particular instrument. The use of a surrogate measure, however, requires the demonstration that scores on that surrogate measure are correlated with the amount of distortion on the personality inventory. If our interest is in measuring the amount of distortion in responses to the personality inventory, it would seem that a direct measure of that distortion is preferable to one that depends on scores from a separate instrument.

Difference Score approaches. A second line of research on response distortion has focused on differences in responses to personality tests from participants responding under different instructional sets. Typically, one instructional set is designed to elicit as little distortion as possible while the other instructional set is designed, either by instruction or incentive, to elicit distortion. To achieve the first instructional set, participants are instructed to respond honestly or told that their responses will have no positive or negative consequences. To elicit distortion, two methods predominate. In some studies participants have been instructed to respond in a fashion that would increase their chances of getting employment, that is, to “fake good”. In other studies, participants have either been given incentives or other indication that

positive consequences will result from better scores on the personality tests or they have been actual applicants for positions (e.g., Rosse, Stecher, Miller, & Levin, 1998)

In these *two-condition* studies, differences in responses between the two conditions have been used as direct measures of the existence or amount of response distortion. When the same participants have been used in the two conditions, the simple difference in scale scores between the two conditions has often served as a measure of response distortion (e.g., McFarland & Ryan, 2000). Although difference scores can serve as reliable measures, their use in exploring the factors related to response distortion is limited because difference scores are typically positively correlated with the subtracting variable in the difference and negatively correlated with the subtracted variable. If the dimension on which the difference is taken is a potential predictor or sequel of response distortion, it will be difficult to separate causal relationships from those resulting from the mathematics of differencing (see Edwards, 1995; 2002 for a detailed discussion). Attempts to circumvent this problem have involved excluding the dimension on which the difference is taken when examining relationships of the difference to personality dimensions. For example McFarland and Ryan (2000) computed difference scores for each of seven personality dimensions and then correlated each difference score variable with only the six other variables that were not part of the dimension on which the difference scores were computed.

Faking has also been modeled as mean differences in scale scores between groups of respondents with different levels of motivation, for example, applicant vs. students or incumbents in between-subjects studies (e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Hough, 1998). A criticism of modeling faking as mean differences of scale scores in between-subjects studies, *inter alia*, is that such between-subjects studies inherently have a low power to detect faking (see Hunter & Schmidt, 1990 for a detailed discussion). Moreover, such studies provided no way to measure individual differences in response distortion and thus no way to examine correlates of those differences.

The use of two-condition research involving an honest condition and one in which participants have been instructed to distort their responses has been criticized by some. It has been argued that the honest condition is one that would rarely if ever be found in real applicant situations. Critics have also argued that instructions to “fake good” may create a mind-set that is unrepresentative of the mind sets of real applicants. Thus, although the two-condition paradigm is one that allows unequivocal estimation of whatever differences exist between the two conditions, it may be that those differences do not reflect real-life applicant response distortion. Clearly a way of estimating response distortion in a single, applicant, situation would be desirable.

Factor analytic approaches. Some research has been directed at examining the factor structure of personality inventories using factor analytic techniques. Schmit and Ryan (1993) factor analyzed responses to individual items of the NEO FFI (Costa & McCrae, 1989) of applicant and non-applicant samples. In the non-applicant sample, they found the expected five-factor solution. However, in the applicant sample, they found that a six-factor solution fit the data best. They found that the sixth factor shared cross-loading with four of the Big 5 dimensions. They labeled the sixth factor an “ideal employee” factor. Later studies (e.g., Frei,

1998; Frei, Griffith, Snell, McDaniel, & Douglas, 1997), using a multi-group CFA approach comparing variance structure of faking good vs. honest groups, showed differences in the number of latent variables, error variances, and intercorrelations among latent variables across groups. All in all, this line of research suggests that faking impacts the measurement properties of personality tests.

A Structural Equation Model of Response Distortion. As mentioned earlier, Biderman and Nguyen (2004) proposed a SEM model wherein faking is conceptualized as an individual difference construct. This model as originally presented was based on the simple notion that responses to a personality scale item under honest instructions are a function of the respondent's position on whatever dimension the items represent and random error. Letting Y represent the observed score on an instrument measuring a personality dimension, D represent the respondent's true position on the dimension and E represent random error, this conceptualization is expressed simply as

$$Y = D + E.$$

In situations in which respondents are induced or instructed to fake, the score on an item is conceptualized as being the sum of the respondent's position on the dimension of interest, D , and the amount of distortion or faking characteristic of that respondent, F . Thus, under faking conditions,

$$Y = D + F + E.$$

These conceptualizations translate into the path diagram presented in Figure 1 for a single dimension in the two-condition paradigm. In the diagram, TH1, TH2, and TH3 are scores on three indicators of the dimension obtained under instructions to respond honestly, and TF1, TF2, and TF3 are scores on the same or equivalent indicators obtained under conditions of inducement or instruction to fake. EH1, EH2, and EH3 are residual latent variables as are EF1, EF2, and EF3.

 Insert Figure 1 about here

The above model was generalized by Biderman and Nguyen (2004) to model applicant faking of multiple dimensions/variables by assuming that the same tendency to distort (F) is applicable to all responses observed in the faking conditions. For example, in application of the model to a questionnaire containing items measuring the Big 5 personality dimensions, there would be five latent variables, each representing one of the Big 5 dimensions and a sixth latent variable representing tendency to distort. This application is illustrated in Figure 2.

 Insert Figure 2 about here

As shown in Figure 2, applicant faking is viewed as a type of common method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Biderman and Nguyen (2004) proposed that the latent variable, F, representing amount of response distortion would be indicated only by scores from the condition in which participants were instructed to fake. However, since there is much evidence that common method variance is applicable in self-report data without instructions or incentives to fake, it appears that a natural modification of their model would be one in which a common method factor indicated by the honest condition observed scores was included. Figure 3 shows this modification to the model of Biderman and Nguyen (2004).

Insert Figure 3 about here

The model in Figure 3 is an example of Case 4A of approaches to modeling multi-trait multi-method data described by Podsakoff et. al. (2003, p. 896). As shown in Figure 3, there are five traits representing the five personality dimensions and two “methods” represented by the two instructional conditions. For this reason, the models including the M variable will be referred to as the MTMM model from now on.

It might be argued that whatever method variance applicable to the honest condition would also be applicable in the faking condition, in addition to whatever extra variance was attributable to response distortion in the faking condition and that all observed variables, honest and faked, should load on the Method (M) latent variable shown in Figure 3. Since the two-condition paradigm includes two distinct experimental conditions, it seemed appropriate for the present time to treat the response distortion in the honest condition as a construct applicable only to the honest condition and distinct from that in the faking condition. Thus we have chosen a view that the instructions or incentives to fake that distinguish the faking condition from the honest condition create a response culture in which tendencies applicable to the honest condition disappear in favor of or are overshadowed by tendencies induced by the faking manipulation.

The present study

In this paper we will consider the application of the MTMM version of the Biderman and Nguyen (2004) model to three different datasets, all of which involved a condition in which respondents were instructed to respond honestly and one or more conditions with incentives or instructions to distort responses. We will examine the extent to which inclusion of a faking latent variable indicated by scores in faking conditions and of a method variance latent variable indicated by scores in honest conditions contribute to the fit of the model. Finally we will present preliminary data on the feasibility of estimating response distortion from only the condition, one in which incentives or instructions to fake are present.

METHOD

The data to which the model was applied were gathered in three separate investigations. The first (Biderman & Nguyen, 2004; Nguyen, Biderman, & McDaniel, 2005) involved a 50-item Big 5 questionnaire based on the Goldberg IPIP items (Goldberg, Johnson, Eber, Hogen,

Ashton, Cloninger, & Gough, 2006) and a situational judgment test, the Work Judgment Survey described by Smith and McDaniel (1998). Respondents (N=203) filled out the Big 5 and SJT under instructions to respond honestly and the same questionnaire again under instructions to “fake good”. Order of experience of the honest and fake conditions was counterbalanced. Participants responded to the SJT by indicating which course of action would be the best and worst to take and also which course of action they would be most likely to take and which they would be least likely to take. Only the most likely/least likely responses were analyzed in the present study. Participants also completed the Wonderlic Personnel Inventory (Wonderlic, 1999). A detailed description of the method for this study can be found in Nguyen et al. (2005).

The second dataset to be reviewed here (Wrensen & Biderman, 2005) involved only the Big 5 questionnaire mentioned above. Participants were college undergraduates (N= 173) who filled out the same 50 item Big 5 questionnaire used in the first study under instructions to respond honestly and again under instructions to “fake good”. As above, order of experience of the conditions was counterbalanced. Participants also completed the Wonderlic Personnel Inventory and several other questionnaires under instructions to respond honestly. The data of those other questionnaires will not be analyzed here.

The third dataset (Clark & Biderman, 2006) involved Goldberg’s 100 item Big 5 scale, the Paulhus Self-Deception scale, the Paulhus Impression Management scale, and a Product Familiarity scale developed especially for the project. For the product familiarity scale, respondents were given names of products of technology and asked to rate their familiarity with those products. Each of the scales was broken into three alternative forms. Undergraduates (N=168) were given a different form of each scale under three conditions. In the first condition, participants were instructed to respond honestly. In the second, they were told that the persons receiving the top three scores would receive a \$50 gift certificate to a local mall although they were reminded to respond honestly. In the third condition, participants were instructed to “fake good”. Because of the possibility of carry over from the incentive condition, the second condition described here, the order of experience of the three conditions was not counterbalanced. All participants received them in the same order: honest followed by incentive followed by instructed faking.

For all datasets, three testlets were used as indicators of each dimension. It was felt that testlets would provide indicators that were more nearly continuously and normally distributed than would individual items (Bandalos, 2002). Each testlet was formed by averaging the responses to individual items. For the first two studies, each Big 5 testlet was the average of three Big 5 items. (Of the 10 IPIP items available for each dimension, the item with the lowest communality in an exploratory factor analysis of the items for that dimension was not used.) For the SJT measure from the first study, each testlet was the average of 10 individual items. The same testlets were used twice – once under instructions to respond honestly and again under instructions to fake good. For the last study, each testlet was the average of two items. For this study, different testlets were used in each condition.

The MTMM model described above was applied to the data of each study. This model included a latent variable representing each of the dimensions included in the study. It also included a latent variable representing the faking aspect of the study. Because participants were

instructed to fake in the first two studies; any response distortion or faking observed would best be conceptualized as due to faking ability, and these latent variables have been labeled FA in the figures describing these studies. The third, Clark and Biderman (2006) study, included two faking conditions. In one condition an incentive to fake was present. For this condition, any response distortion or faking would best be conceptualized as due to the propensity or motivation to fake. For this reason, the latent variable indicated by testlets from this condition is labeled FP, for faking propensity. In the last condition of the Clark and Biderman (2006) study, participants were instructed to fake, and the latent variable representing that condition is labeled FA, as in the first two studies.

Finally, the model applied to each dataset included a latent variable representing method variance. That variable was indicated by the testlets from the honest condition of each study. In all models that latent variable is labeled M.

After applying the MTMM model to the two-condition (three in the case of the Clark and Biderman (2006) study), the models were applied to the data of each condition separately – to the honest condition data of each study and to the instructed faking condition of each study, and, in the last study, to the incentive condition data. Factor scores from each application were computed and correlated with factor scores of corresponding latent variables (M, FA, or FP) computed from the application of the two and three-condition models.

All models were applied using Amos Version 5.0 and MPlus 3.1 (Muthen & Muthen, 1998-2004).

RESULTS

Figure 4 presents the Amos path diagram of the original Biderman and Nguyen (2004) model applied to the first dataset. In the development of the model, it was found that fit was significantly improved if the latent error variables of the faking instruction testlets were allowed to be correlated. The need for such correlated error terms may reflect dimension specific covariance in faking ability. An alternative way to address such covariance would be to include a separate faking latent variable for each dimension. It was decided that allowing the error latent variables to covary represented a slightly more parsimonious solution.

Insert Figure 4 about here

In addition to correlated errors among the faking testlets, significant improvements in goodness of fit were obtained by allowing the error latent variables for corresponding honest and faking testlets to covary.

Figure 5 presents the application of the MTMM version of the model to the Biderman and Nguyen (2004) dataset. Table 1 presents goodness of fit statistics for the application. The overall chi-square was 779.2, $p < .001$, with CFI = .947 and RMSEA = .049. The table also presents chi-square difference statistics comparing the fit of the MTMM model with two separate special models – one without a faking ability latent variable and a second without a method

variance latent variable. Dropping the faking ability latent variable increased the chi-square by 391.2 ($p < .001$) while dropping the method variance latent variable increased it by 97.4 ($p < .001$). This suggests that goodness of fit was improved by inclusion of both, particularly the faking ability variable.

Insert Figure 5 about here

Figure 6 presents the MTMM model applied to the data of Wrensen and Biderman (2005). Table 1 presents goodness of fit statistics. As was the case for the Biderman and Nguyen (2004) data, the fit of the model is adequate when the CFI and RMSEA are considered although the chi-square statistics suggests the need for improvement. Again, the chi-square goodness of fit statistic was increased significantly by dropping the faking ability latent variable ($\Delta\chi^2 = 345.8, p < .001$) and the method variance latent variable ($\Delta\chi^2 = 43.9, p < .001$).

Insert Figure 6 about here

Figure 7 presents the Amos path diagram of the MTMM model applied to the data of Clark and Biderman (2006). Recall that this experiment had three conditions – an honest response condition, an incentive condition, and an instructed faking condition. Moreover, the dataset included three measures in addition to the Big 5 – a self-deception, an impression management, and a product familiarity meaSURE. Finally, this dataset is different from the previous ones in that alternative forms of the instruments were used in the three conditions so whereas the design was a within-subjects design, it did not involve repeated measures of identical testlets across the three conditions. For this reason, there were no “across-condition” covariances estimated as was the case with the previous two datasets.

Insert Figure 7 about here

As can be seen in Figure 7, the goodness of fit statistics are considerably worse than in the application to the previous two datasets. It is presumptuous to attempt to attribute the poor goodness of fit to only idiosyncratic unmodeled covariances among the testlets. However, at the present time, we have no explanation for the lack of fit. We note that Clark and Biderman (2006) applied a model without a method variance latent variable to whole-scale scores rather than testlets and found acceptable fit of that model to the whole-scale data. The testlets used in the applications described here were formed arbitrarily, and we have not yet had time to examine their equivalence or to consider possible wording similarities/differences between the testlets.

In post hoc examinations of the fit of the model to various subsets of the Clark and Biderman (2006) data, we discovered that restricting the analyses to only the Big 5 testlets resulted in substantial improvement to the fit statistics. Since the Big 5 dimensions are common

to the other datasets observed here and are ones that would be most likely to be of interest to those investigating personality variables in selection, we decided to exclude the self deception, impression management, and product familiarity scales from the analyses reported here. Figure 8 presents the MTMM model applied to the Big 5 testlets from the Clark and Biderman (2006) dataset. Since this dataset included three conditions - an honest responding condition, an incentive condition (indicated by D, for “dollar” in Figure 8) and an instructed faking condition - three latent variables were included in the MTMM model. The latent variable, FP, for faking propensity, was indicated by testlets from the incentive (“dollar”) condition.

With the exception of the chi-square statistic, goodness of fit of the MTMM model to the Big 5 data was better than fit to the data of all the measures. Dropping faking ability latent variable results in an increase in chi-square of 328.3 ($p < .001$). Dropping the faking propensity latent variable also resulted in a significant increase ($\Delta\chi^2 = 101.2, p < .001$). But dropping the method variance latent variable resulted in only a negligible increase ($\Delta\chi^2 = 21.9, p > .05$).

 Insert Figure 8 about here

Estimates of key parameters.

Table 2 presents estimates of key parameters from the above models. We would expect the variances of all the latent variables to be significantly greater than zero for all three models. In general this was found in all three applications, although the evidence for the Clark and Biderman (2006) data is the weakest.

For the first two datasets correlations among the Big 5 latent variables were generally close to zero, not surprising given the essentially orthogonal nature of the dimensions. However the correlations among the latent variables representing the Big 5 dimensions were all positive for the Clark and Biderman (2006) dataset, suggesting a possible misspecification in the model for those data. Note that the models were applied with the restriction that the method/faking latent variables were orthogonal to the dimension latent variables. Results related to the relaxation of this assumption are presented later.

The method variance latent variable was positively correlated with the faking ability latent variable in the first two studies. In the 3rd study, none of the method/faking latent variables was significantly correlated with any other.

Loadings of testlets onto their respective dimension latent variables were positive. Loadings of testlets from the honest condition were generally the largest. Those from the instructed faking conditions in all three datasets were much smaller than the loadings from the honest conditions. Those from the incentive condition in the Clark and Biderman (2006) dataset were just slightly smaller than those from the honest condition.

Loadings of the faking condition testlets onto the faking latent variables were always positive and sometimes nearly as large as the loadings of honest testlets onto the latent trait

variables. While smaller, loadings of testlets onto the method variance latent variable were positive in the first two studies. However, for the Clark and Biderman (2006) data, honest conditions testlets loadings onto the method variance latent variable were small with about as many negative as positive, in keeping with the small change in the chi-square goodness of fit statistic associated with that latent variable.

Construct Validity.

Each study involved administration of scales other than those involved in the faking administration. It was decided that for this review, with the exception of the SJT test from the Biderman and Nguyen (2004) study, analyses involving construct validity would only involve those variables common to all three studies. For each dataset, an analysis was conducted in which all of the Big 5 latent variables and the Wonderlic and the SJT latent variable in the case of the Biderman and Nguyen (2004) study were allowed to correlate with the method variance and faking latent variables.

The most consistent finding across the three studies is that the correlation of the Wonderlic with faking ability was significantly positive in the first two studies and approached significance in the third. The faking ability latent variable was not significantly related to any of the Big 5 latent variables in any of the datasets. The faking propensity latent variable from the Clark and Biderman (2006) dataset also was not related to any of the Big 5 latent variables.

One-condition estimates.

For each dataset, a CFA was performed on the data of each individual instructional condition. That is, a CFA was performed on only the honest testlets, then a separate CFA was performed on only the faking testlets. In the third, Clark and Biderman's (2006) dataset, a CFA was performed on the instructed faking condition testlets and a separate CFA was performed on the incentive condition testlets. For each CFA, a method or faking latent variable was estimated along with the latent variables appropriate for the study. Figures 9 through 11 present path diagrams of the models estimating faking ability for each dataset. Figure 12 presents the path diagram of the model estimating faking propensity from the Clark and Biderman (2006) study. The path diagrams of the models estimating method variance are similar.

Insert Figures 9 thru 12 about here

Factor score coefficients were obtained from each analysis, and were used to compute factor scores for each method/faking factor. Factor score coefficients were also obtained for the method and faking latent variables from the application of the model to two- and three-condition data described earlier – specifically from the models illustrated in Figures 5-7 – and used to generate factor scores for the method and faking factors.

Correlations between the one-condition factor scores and two-condition factors scores from the same dataset were then computed. Table 4 presents determinacy coefficients for each

set of factor scores and correlations between the corresponding method and faking factors for the three datasets. Inspection of the table shows that the correlations between the one- and two-condition method factors from each dataset were quite low. On the other hand, the correlations between the one- and two-condition faking ability factors were larger than .9 for each dataset. The correlation between the one- and two-condition faking propensity factor from the Clark and Biderman (2006) dataset was slightly larger than .5. Figure 16 presents scatterplots of the faking latent variable factor scores.

Insert Figure 16 about here

DISCUSSION

In this paper, we reviewed the use of structural equation models to measure applicant response tendencies that might affect personality test scores. Our analyses strongly suggest that response tendencies that might be best characterized as faking or distortion affected test scores completed under faking induced conditions. Our analyses suggest that a response tendency that might be characterized as method variance may affect scores of personality tests completed under conditions conducive to responding honestly. Finally, very high correlations were obtained between factor scores of faking ability from two-condition data and factor scores from one-condition data, suggesting that it may be feasible in some instances to estimate faking ability from a single condition. The evidence is less persuasive for the estimation of faking propensity from a single condition.

The chi-square statistic indicated significant deviations from acceptable fit in all datasets, as is common in application of structural equation models to large datasets. Although we have not yet had the opportunity to completely explore the reasons for the lack of fit shown by the chi-square statistics, it is our assumption that this is due primarily to unmodeled idiosyncratic correlations among testlets due to patterns of wording and scoring.

The applications of the models presented here necessarily are confirmatory factor analyses of MTMM data. CFAs applied to such data are notoriously prone to failure to converge. We encountered some convergence problems in the applications described here. For example, it was found that the error latent variables for one of the dimensions could not be allowed to be correlated in application of the model to the Clark and Biderman (2006) data (See Figure 7). We also found one Heywood case in which the variance of a residual latent variable was estimated as a negative value.

We were particularly concerned about the robustness of the application of the model to one-condition data. For example, the use of faking latent variable factor scores in the selection process would not be feasible if nonconvergence were a problem. In an effort to examine the robustness of the application of the model to one-condition data, simulated one-condition data were generated using the Monte Carlo feature of Mplus. Table 5 presents the results of those simulations. As can be seen from inspecting the table, when data were generated with equal loadings of testlets on all factors, the proportion of analyses that converged was quite small.

However, the convergence rate was increased considerably by generating data in which the loadings of the three testlets on each factor were varied from one testlet to another. Such a result would be expected based on the work of Kenny and Kashy (1992) who showed that some MTMM models are not identified when loadings are equal. Clearly further exploration of the conditions under which convergence occurs is needed.

We note that the loadings of testlets onto the method variance latent variable were generally much smaller than those on the faking ability or faking propensity latent variables. This suggests that method variance may not be as important in determining responses to testlets in the honest condition as faking ability or faking propensity in the faking conditions. Moreover, the small correlations of the two-condition method variance factor scores with the one-condition method variance factor scores suggest that the method variance is not well-captured by the models presented here.

Unfortunately, there was also a fairly low correlation between Clark and Biderman three-condition faking propensity factor scores and their one-condition counterparts. This suggests that whatever individual differences represented by the three-condition factor scores were not completely captured by the one-condition scores. At first glance, this does not bode well for the usefulness of this approach in an applicant setting, where it would be expected that faking propensity would be a major source of individual differences in response distortion. However, there are two reasons to suspend judgment on this issue at the present time. First, the Clark and Biderman (2006) study used testlets composed of only two items, rather than three. Thus, those testlets were less reliable than the three-item testlets employed in the other two studies. Unreliability of the indicators probably leads to less stable solutions in applications of models such as this. Secondly, the incentive manipulation used by Clark and Biderman was not particularly forceful. It is possible, as is suggested by the differences in loadings between the method variance and faking latent variables, that the strength of manipulation affects the stability of estimates of the parameters of the model. This speculation is supported by the lower determinacy values for the faking propensity factor as shown in Table 4.

The issue of the generality of response distortion when multiple personality dimensions are assessed has not been fully explored. In studies employing surrogates of response distortion, such as social desirability, it was seemingly assumed that SD was a characteristic applicable to a variety of instruments. In two-condition studies, some investigators have computed difference scores on multiple dimensions and computed a single measure of faking. For example, Mersman & Shultz (1998) assessed the Big 5 and computed the average of the difference scores on the 5 dimensions as an overall measure of faking. Other studies have computed separate faking estimates for different dimensions. For example, McFarland and Ryan (2000) computed difference scores on seven separate dimensions and did not compute a single overall estimate of faking ability. Thus, whereas the notion of the tendency to distort responses being applicable across testing instruments was implicit in applicant faking research using surrogate measures such as social desirability, the conceptualization of a general faking tendency has been eschewed by many researchers using two-condition data. The generally positive loadings of all testlets on the single faking ability latent variable in all three studies and on the faking propensity latent variable in the third study presented here strongly suggest that both the propensity and ability to distort responses are general phenomena across instruments.

We propose that the faking propensity factor identified in the present study may be the same factor labeled an “ideal employee” factor in the Schmit and Ryan (1993) study. Schmit and Ryan used an exploratory factor analysis of the applicant data in which their sixth factor was identified. The sixth factor identified by Schmit and Ryan (1993) did not load on all the observed scores in their EFA. We consider a factor with positive loadings on all the five dimensions as a better example of an ideal employee factor than one with the loading pattern found by Schmit and Ryan. Furthermore, we view the connotations of “ideal employee” as being consistent with “faking good”. Thus our view is that the model and applications presented here represent an extension and perhaps clarification of the research of Schmit and Ryan (1993).

The model presented here has been presented as a CFA model for MTMM data. However, the model has the same form as a two time-period longitudinal growth model for multiple factors. Referring to Figure 1, the D latent variable plays the role of the intercept latent variable in a LGM while the F latent variable plays the role of the slope latent variable (e.g., Duncan, Duncan, Strycker, Li, & Alpert, 1999). In Figure 2, the extension to multiple dimensions, the different dimensions represent different intercepts of different growth processes while the single F latent variable represents a common slope. Obviously, multiple slope latent variables – one for each process - could have been included in the model. Figure 14 presents such a model as applied to the Biderman and Nguyen (2004) dataset. This model represents the common faking ability as a higher order factor indicated by individual dimension faking ability factors. This variation of the model might be considered a factor of curves model, albeit a limiting case of such a model with only two time points per “curve” (McArdle, 1988). Preliminary application of a multiple faking factor two-wave LGM to the Biderman and Nguyen (2004) data suggest that the alternative conceptualization fits the data as well as the model presented here. However, that model requires a second order factor to represent a common faking effect across dimensions. For that reason, we felt the MTMM model presented here was more parsimonious.

Conceptualizations of two time-period data such as that in Figure 1 form the core of LGM models and specific presentations to two-wave data have preceded the conceptualization presented here. Embretson (1991) presented a model of change in ability across time that is similar to the model of the Biderman and Clark (2005) data. Raykov’s (1992) two-indicator structural model for base-free measurement of change is similar to that presented Figure 1. Cribbie and Jamieson (2000) presented a model for measuring correlates of change that is similar to the model presented in Figure 1.

The similarity of the models applied here to LGM models raises the issue of estimation of means and intercepts. Such an expanded model could be applied to the data reviewed here. In fact, Clark and Biderman (2006) include such an expanded model. However, our focus on faking ability and faking propensity as individual difference variables obviated the need for estimation of means and intercepts. Clearly in applications focused on comparing mean amounts of faking between groups or between conditions, applications in which means and intercepts are estimated will be required.

Two characteristics of the model represented in Figure 1 should be noted. First, it assumes that responses to personality items are additive functions of the respondents' self-reported positions on those dimensions and the respondents' tendencies to distort. Second, implicit in this assumption is an assumed process: Respondents assess their own agreement with the item (presumably partly determined by their position on the dimension) and then distort that agreement by an amount represented by their overall tendency to distort. The strong evidence for individual differences in faking ability and propensity from the applications of the model presented here are consistent with this assumption. However, it may be that respondents adopt other strategies when given incentives or instructions to fake. For example, under faking conditions, it may be that some respondents target specific responses while completing an instrument. Such targeted responding certainly does not correspond to the assumptions of the models presented here and might account to some extent for the poor fit of the models as indicated by the chi-square statistics.

The separate estimation of dimension and faking latent variables shown here would not have been possible without the use of multiple indicators for each dimension. We used three indicators for each dimension. We guess that the use of more indicators per dimension would result in greater stability of estimates of all the parameters of the models as long as those indicators possessed appropriate distributional characteristics. We note that the use of multiple indicators for the Big 5 dimensions would have been more difficult and expensive were it not for the ready availability of the IPIP items (Goldberg, Johnson, Eber, Hogen, Ashton, Cloninger, & Gough, 2006).

The admittedly sketchy results concerning construct validity of the faking ability and faking propensity factors presented here leaves much room for further research. We are comforted by the consistent positive correlations of the faking ability latent variable with cognitive ability. This is in agreement with our own conceptualization of faking ability as an ability like other abilities, one that is carried out more thoroughly by persons high in cognitive ability.

Only one of the three studies measured what many people consider to be the essence of faking behavior – what we have called faking propensity. Since the instructed faking paradigm creates a situation in which participants are given permission to fake, we agree with critics that this paradigm does not engender the same tendencies that are probably present in applicant populations or others with an incentive to distort but without permission to do so. Clearly, more research involving faking incentive designs is needed to discover the nomological net of constructs related to faking propensity as it is defined using this structural equation model.

We investigated the relationship of the faking latent variables to latent variables representing all the Big 5 dimensions simultaneously. The validity of such a test, given that the faking ability latent variable is intimately tied to the dimension latent variables, depends on the adequacy of the model of all the behavior in the experimental situation. Examining the relationship of faking to the dimension scores is analogous to examining the relationship of the slope parameter to the intercept parameter in longitudinal growth models. Misspecification of the mode of growth, for example, neglecting to account for a nonlinear trend, may impose a correlation between the two parameters, confounding attempts to examine relationship between

the two. For this reason, we approached the analyses reported in Table 4 with a certain amount of trepidation, and consider the results of the analyses reported there to be tentative at best. We will feel more confident in results such as those as our confidence in the adequacy of the models of faking being developed here increases. We note that the lack of significant correlations of the method and faking latent variables with the Big 5 dimensions is quite different from what would certainly have been found had the faking effects been defined as simple difference scores.

It's probably safe to say that most of the past research on MTMM models has focused on estimating method effects in an effort to be rid of them or to partial out their effects. Our treatment of the M latent variable here was similar. We had no theoretical interest in it beyond treating it as a nuisance variable whose effects were to be eliminated. Our view of the FA and FP latent variables is far different, however. We view both as substantive constructs of some theoretical interest. This is particularly true in the case of the FP latent variable. Other studies focusing on the theoretical nature of method effects have typically utilized direct measures of the construct presumed to underlie the effects, such as social desirability or negative affectivity (e.g., Jex & Spector, 1996). However studies examining theoretical relationships involving method variance measured as a general factor sharing all indicators with other trait factors are much less common (but see Quilty, Oakman, & Risko, 2006). Thus the applications of the model presented here represent somewhat of a departure from the traditional analysis of MTMM data in that the primary interest of the research is the relationships of other variables to the FA and FP variables.

We were disappointed to find such a low correlation between the FP factor scores estimated from the multi-condition analysis of the Clark and Biderman (2006) data and those estimated from the incentive condition only. If such factor scores were to be used as measures of faking propensity in an applicant sample, for example, it would have to be demonstrated that the one-condition estimates were highly correlated with estimates taken from data that included an honest response condition. More research is needed to determine the boundary conditions for adequate estimates. A study with a stronger incentives than that used by Clark and Biderman (2006) would be a first step.

Further research is also need to clarify the conditions under which the one-condition model will yield any estimates at all. The exploratory simulation conducted here suggested that having items or testlets with identical loadings on a factor may decrease the likelihood of achieving a solution. This discovery was surprising, and should be explored further.

REFERENCES¹

- 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.
- Chan, D. (1998). The conceptualization and analysis of change over time: an integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421-483.
- Clark, J. M. & Biderman, M. D. (2006). A Structural Equation Model Measuring Faking Propensity and Faking Ability. Paper accepted for presentation at the 21st Annual Conference of The Society for Industrial and Organizational Psychology, Dallas, TX, May.
- Costa, P.T., Jr., & McCrae, R.R. (1989). *The NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Duncan, T.E., Duncan, S.C., Strycker, L.A., Li, F., & Alpert, A. (1999). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues and Applications*. Mahwah, NJ: Erlbaum.
- 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.
- Ellingson, J.E., Sackett, P.R., & Hough, L.M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.
- Embretson, S. E. (1991). Implications of a multidimensional latent trait model for measuring change. In Collins, L.M. & Horn, J.L. (Eds). *Best methods for the analysis of change: recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.
- Frei, R.L. (1998). *Fake this test! Do you have the ability to raise your score on a service orientation inventory*. University of Akron. Unpublished doctoral dissertation.
- Frei, R.L., Griffith, R.L., Snell, A.F., McDaniel, M.A., & Douglas, E.F. (1997). *Faking of non-cognitive measures: Factor invariance using multiple groups LISREL*. Paper presented at the 12th Annual Meeting of the Society for Industrial & Organizational Psychology: St. Louis, MO.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 50*, 84-96.
- 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.
- Hunter, J.E., & Schmidt, F. L. (1990). *Methods of Meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.

- Jex, S. M., & Spector, P. E. (1996). The impact of negative affectivity on stressor strain relations: A replication and extension. *Work and Stress, 10*, 36-45.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*, 165-172.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattell & J. Nesselrode (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561-614). New York: Plenum Press.
- McFarland, L.A., & Ryan, A. M., (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812-821.
- Muthen, L. K., Muthen, B. O. (1998-2004). *Mplus User's Guide*. 3rd Ed. Los Angeles, CA: Muthen & Muthen.
- Nguyen, N.T., & McDaniel, M.A. (2001). *The influence of impression management on organizational outcomes: A meta-analysis*. Poster presented at the 16th Annual Meeting of the Society for Industrial & Organizational Psychology. San Diego, CA.
- Nguyen, N.T., Biderman, M.D., & McDaniel, M.A. (2005). The effect of Response Instructions on faking a Situational Judgment Test. *International Journal of Selection and Assessment, 13*, 250-260.
- 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.
- Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling, 13*, 99-117.
- Rogers, WM, & Schmitt, N (2004). Parameter recovery and model fit using multidimensional composites: A comparison of four empirical parceling algorithms. *Multivariate Behavioral Research, 39*, 379-412.
- 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.
- 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.
- Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. *Journal of Applied Psychology, 85*, 50-64.
- Wonderlic, Inc. (1999). *Wonderlic's Personnel Test manual and scoring guide*. Chicago: IL: Author.
- Wrensen, L. B., & Biderman, M. D. (2005). Factors related to faking ability: A structural equation model application. Paper presented at the 20th Annual Conference of The Society for Industrial and Organizational Psychology, Los Angeles, CA.

¹ The Biderman & Nguyen (2004) and Wrensen & Biderman (2005) papers are available from www.utc.edu/Michael-Biderman .

Table 1. Goodness of fit statistics

Biderman & Nguyen, 2004

	MTMM	Method Variance	Without Faking Ability
X ² (df)	681.8 ^c	$\Delta X^2=97.4^c$	$\Delta X^2=391.2^c$
df	506	19	19
CFI	.963	.947	.885
RMSEA	.041	.049	.072

Wrensen & Biderman, 2005

	MTMM	Method Variance	Without Faking Ability
X ² (df)	546.7 ^c	$\Delta X^2=43.9^c$	$\Delta X^2=345.8^c$
df	350	16	16
CFI	.939	.939	.845
RMSEA	.057	.057	.091

Clark & Biderman, 2006 – Big 5 testlets only

	MTMM	Method Variance	Without Faking Ability	Without Faking Propensity
X ² (df)	1249.5 ^c	$\Delta X^2=21.9$	$\Delta X^2=328.3^c$	$\Delta X^2=101.2^c$
df	874	17	17	17
CFI	.859	.859	.744	.829
RMSEA	.051	.051	.068	.056

^a p<.05^b p<.01^c p<.001

Table 2. Estimates of key parameters of the MTMM models. Where appropriate, entries are standardized values or means of standardized values. Entries for variances are critical ratios.

Parameter	Study		
	Biderman & Nguyen	Wrensen & Biderman	Clark & Biderman
Critical ratios of variances of latent variables			
σ^2_E	6.64 ^c	6.99 ^c	4.510 ^c
σ^2_A	4.14 ^c	2.36 ^b	3.34 ^c
σ^2_C	4.68 ^c	6.38 ^c	4.19 ^c
σ^2_S	6.32 ^c	5.35 ^c	3.37 ^c
σ^2_O	1.15	5.27 ^c	4.26 ^c
σ^2_M	3.30 ^c	2.90 ^b	1.14
σ^2_{FA}	6.41 ^c	5.85 ^c	1.84
σ^2_{FP}			2.08 ^a
σ^2_{SJT}	4.49 ^c		
Correlations between latent variables			
E<->A	.26 ^b	.17	.44 ^c
E<->C	-.06	.02	.11
E<->S	.15	-.04	.19 ^a
E<->O	.07	.26 ^b	.53 ^c
A<->C	.14	.18	.28 ^a
A<->S	-.17	-.17	.31 ^b
A<->O	-.12	.10	.26 ^a
C<->S	-.06	.27 ^b	.13
C<->O	-.18	.10	.21 ^a
S<->O	-.23	.27 ^b	.20
M<->FA	.59 ^c	.23 ^b	.01
M<->FP			-.03
FA<->FP			.12
Mean Standardized Loadings of testlets onto Dimension latent variables			
	.H FA	. H FA	. H FA FP
E->Honest/E->Faked	.81 / .36	.85 / .23	.69 / .16 / .58
A->Honest/A->Faked	.71 / .30	.56 / .25	.61 / .26 / .52
C->Honest/C->Faked	.61 / .27	.79 / .38	.67 / .18 / .50
S->Honest/S->Faked	.75 / .30	.74 / .42	.68 / .19 / .61
O->Honest/O->Faked	.45 / .13	.71 / .38	.58 / .17 / .57
SJT->Honest/SJT->Faked	.64 / .51		
Mean Standardized Loadings of testlets onto Method and Faking latent variables			
	.M FA	. M FA	. M FA FP
M->E / FA->E / FP->E	.30 / .63	.08 / .55	-.12 / .57 / .26
M->A / FA->A / FP->A	.29 / .55	.04 / .14	-.05 / .50 / .38
M->C / FA->C / FP->C	.52 / .74	.15 / .62	-.14 / .59 / .39
M->S / FA->S / FP->S	.45 / .73	.39 / .59	.26 / .51 / .36
M->O / FA->O / FP->O	.62 / .72	.20 / .51	.04 / .48 / .27
M->SJT / FA->SJT	.34 / .32		

^a p<.05

^b p<.01

^c p<.001

Table 3. Correlations of Faking Ability, Faking Propensity, and Method Variance latent variables with Big 5 latent variables and Wonderlic scores. From top to bottom, values are from Biderman and Nguyen (2004), Wrensen and Biderman (2005) and Clark and Biderman (2006) respectively.

	E	A	C	S	O	Wonderlic
M	-.12	.17	.31	.14	.56	.35 ^c
	-.03	-.36	-.50 ^a	-.91	-.48	-.21
	.36	.42	.21	.66	.43 ^a	.12
FA	-.08	.05	.09	-.18	.11	.47 ^c
	.10	.11	.01	-.02	-.08	.20 ^a
	-.06	.03	.14	-.09	.04	.15
FP	.00	-.20	.18	-.19	.00	-.02

^a $p < .05$

^b $p < .01$

^c $p < .001$

Table 4. Correlations of factor scores from two-condition and one-condition analyses.

Factor Score Determinacies

	Biderman & Nguyen	Wrensen & Biderman	Clark & Biderman
<u>Two condition factor scores</u>			
Method	.906	.778	.786
Faking Ability	.956	.938	.925
Faking Propensity			.829
<u>One condition factor scores</u>			
Method	NA	.808	NA
Faking Ability	.898	.808	.982
Faking Propensity			.767

Correlations between two- or three-condition and one conditions estimates

	One-condition estimates								
	Biderman & Nguyen			Wrensen & Biderman			Clark & Biderman		
	M	FA		M	FA		M	FA	FP
Two-condition M	.17	.59		.32	.22		.28	.01	.02
Two-condition FA	.19	.92		.10	.91		.12	.90	.00
Two-condition FP							.09	.07	.53

Table 5. Preliminary results of Monte Carlo simulations of one-condition data.

Dimension loadings:	1 1 1	1 .8 .5	1 1 1
FA loadings	1 1 1	1 1 1	1 .8 .5
No. successful solutions	39/100	74/100	77/100
Variances	Means of	Means of	Means of
Latent Pop	Estimates	Estimates	Estimates
Var Value			
E .6	-0.0862	0.5390	0.5294
A .1	0.0743	0.0559	0.0541
C .1	0.0224	0.0212	0.0461
S .1	-0.0220	0.0559	0.0401
O .1	-0.0453	0.0301	0.0528
F .2	0.9049	0.2715	0.2815
FA loadings	Means of	Means of	Means of
<u>Testlet</u>	Estimates	Estimates	Estimates
FETL1	1.0000	1.0000	1.0000
FETL2	1.3291	1.0797	0.7731
FETL3	1.2314	1.0891	0.4879
FATL1	1.1712	1.2997	1.0949
FATL2	1.0980	1.2103	0.9375
FATL3	1.1228	1.1181	0.5977
FCTL1	0.9250	1.1789	1.1207
FCTL2	0.9770	1.1153	0.9469
FCTL3	1.0492	1.1102	0.6137
FSTL1	1.1248	1.2028	1.1274
FSTL2	1.1762	1.1473	0.9288
FSTL3	1.2118	1.1071	0.5703
FOTL1	1.0812	1.1616	1.1204
FOTL2	0.9126	1.1113	0.9400
FOTL3	1.0601	1.1466	0.5976

Example of specification of loadings of testlets on Dimensions and on the Faking Ability factor. The following is what was used for the middle column above.

```

e by fetl1@1  fetl2*.8 fetl3*.5;      !Dimension loadings
a by fatl1@1  fatl2*.8 fatl3*.5;
c by fctl1@1  fctl2*.8 fctl3*.5;
s by fstl1@1  fstl2*.8 fstl3*.5;
o by fotl1@1  fotl2*.8 fotl3*.5;
f by                                           !Faking factor loadings
fetl1*1 fetl2*1 fetl3*1
fatl1*1 fatl2*1 fatl3*1
fctl1*1 fctl2*1 fctl3*1
fstl1*1 fstl2*1 fstl3*1
fotl1*1 fotl2*1 fotl3*1;

```

Figure 1. Path diagram showing the relationships of observed scores to latent variables representing a personality dimension (D) and a tendency to distort responses (F). Observed variables TH1, TH2, and TH3 represent items or testlets obtained under instructions to respond honestly. Those labeled TF1, TF2, and TF3 represent items or testlets obtained under instructions or incentives to distort responses. EH1 . . . EF3 are residual latent variables

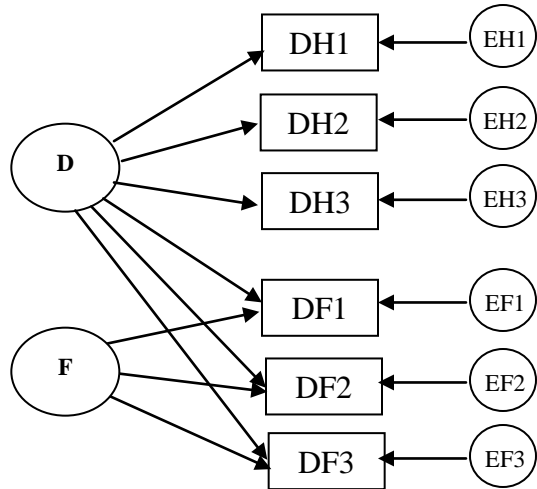


Figure 2. Extension of the model to multiple dimensions as proposed by Biderman and Nguyen (2004). The diagram is based on the application of the model to data of the Big 5 dimensions. The latent variables, E, A, C, S, and O represent Extroversion, Agreeableness, Conscientiousness, Stability, and Openness respectively. The latent variable F represents amount of faking or response distortion. The variables beginning with EH, AH, CH, SH, and OH represent items or testlets observed under instructions to respond honestly. The variables beginning with EF, AF, CF, SF, and OF represent items or testlets observed under instructions or incentive to fake or distort responses. Residual latent variables have been omitted for clarity.

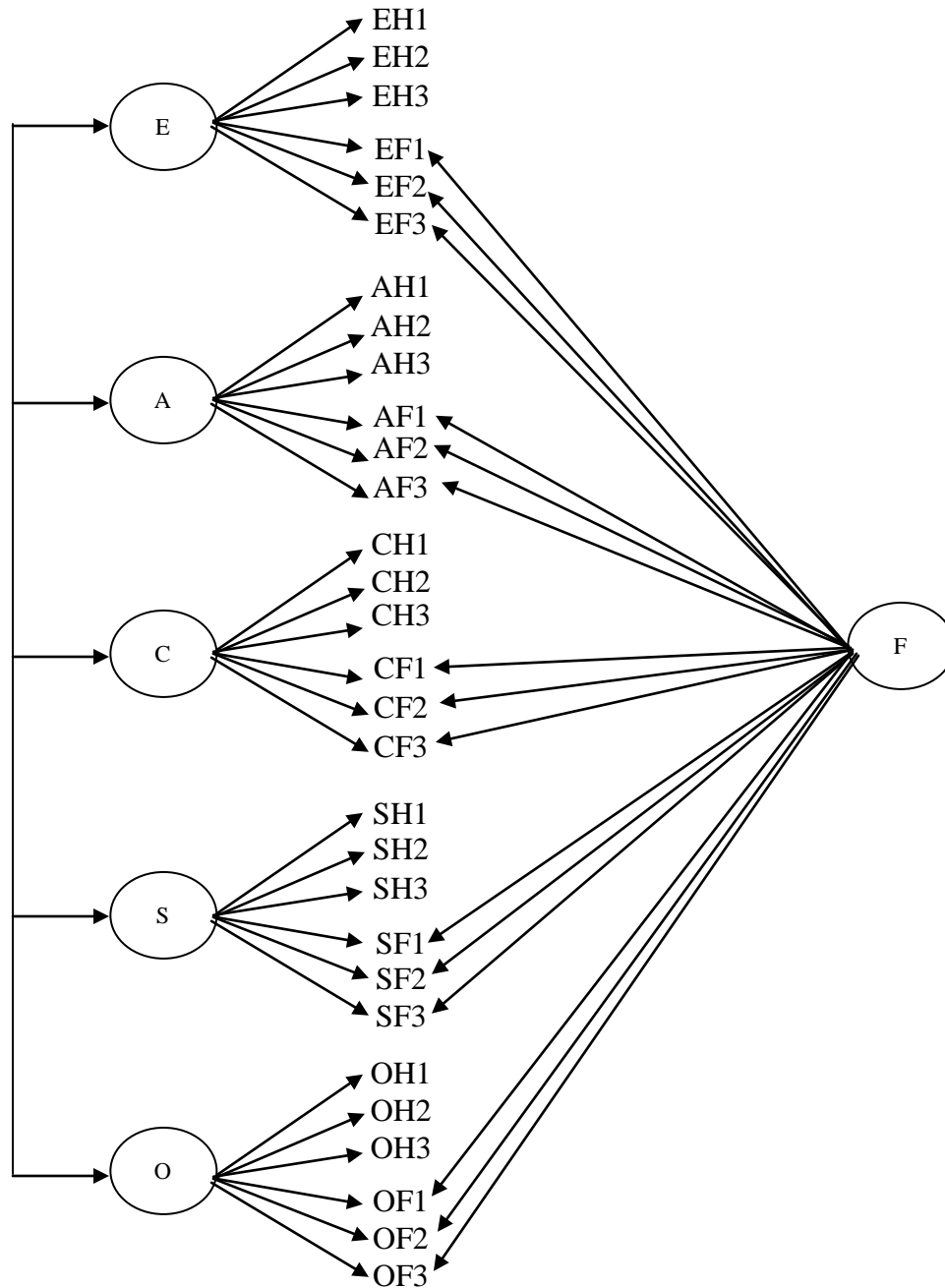


Figure 3. Biderman and Nguyen (2004) model elaborated to include a latent variable representing method variance in the honest response condition. The diagram is based on the application of the model to data of the Big 5 dimensions. The latent variables, E, A, C, S, and O represent Extroversion, Agreeableness, Conscientiousness, Stability, and Openness respectively. The latent variable F represents amount of faking or response distortion. Latent variable M represents amount of method variance. The variables beginning with EH, AH, CH, SH, and OH represent items or testlets observed under instructions to respond honestly. The variables beginning with EF, AF, CF, SF, and OF represent items or testlets observed under instructions or incentive to fake or distort responses. Residual latent variables have been omitted for clarity.

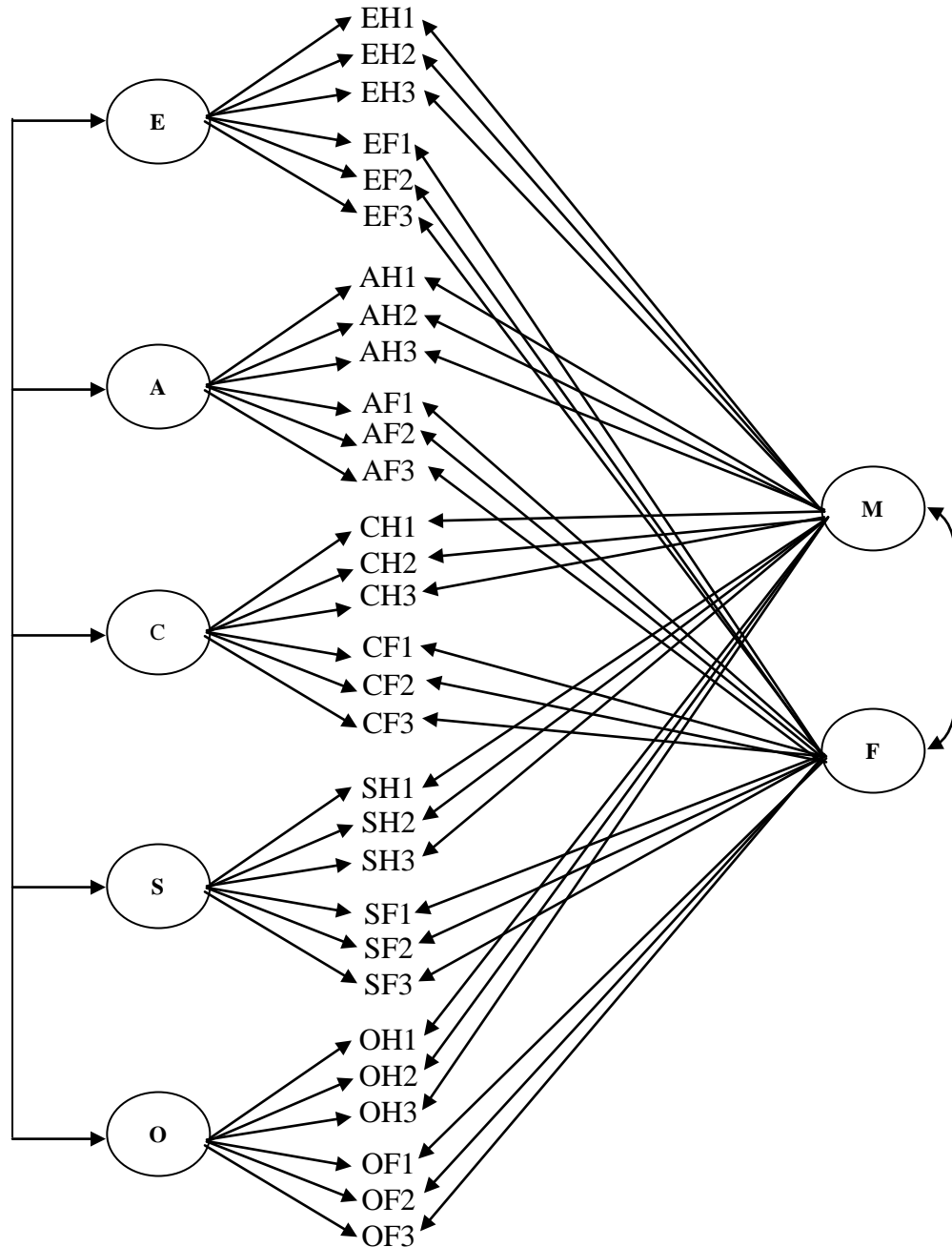


Figure 4. Amos path diagram of original model presented by Biderman and Nguyen (2004). Covariances between faking condition testlets were estimated as were covariances between corresponding testlets from honest and faking conditions.

SIOPM4_MeansNotEstimated 3/18/4
RMSEA = .049
CFI = .947
Chi-square = 779.215
df = 525
p = .000

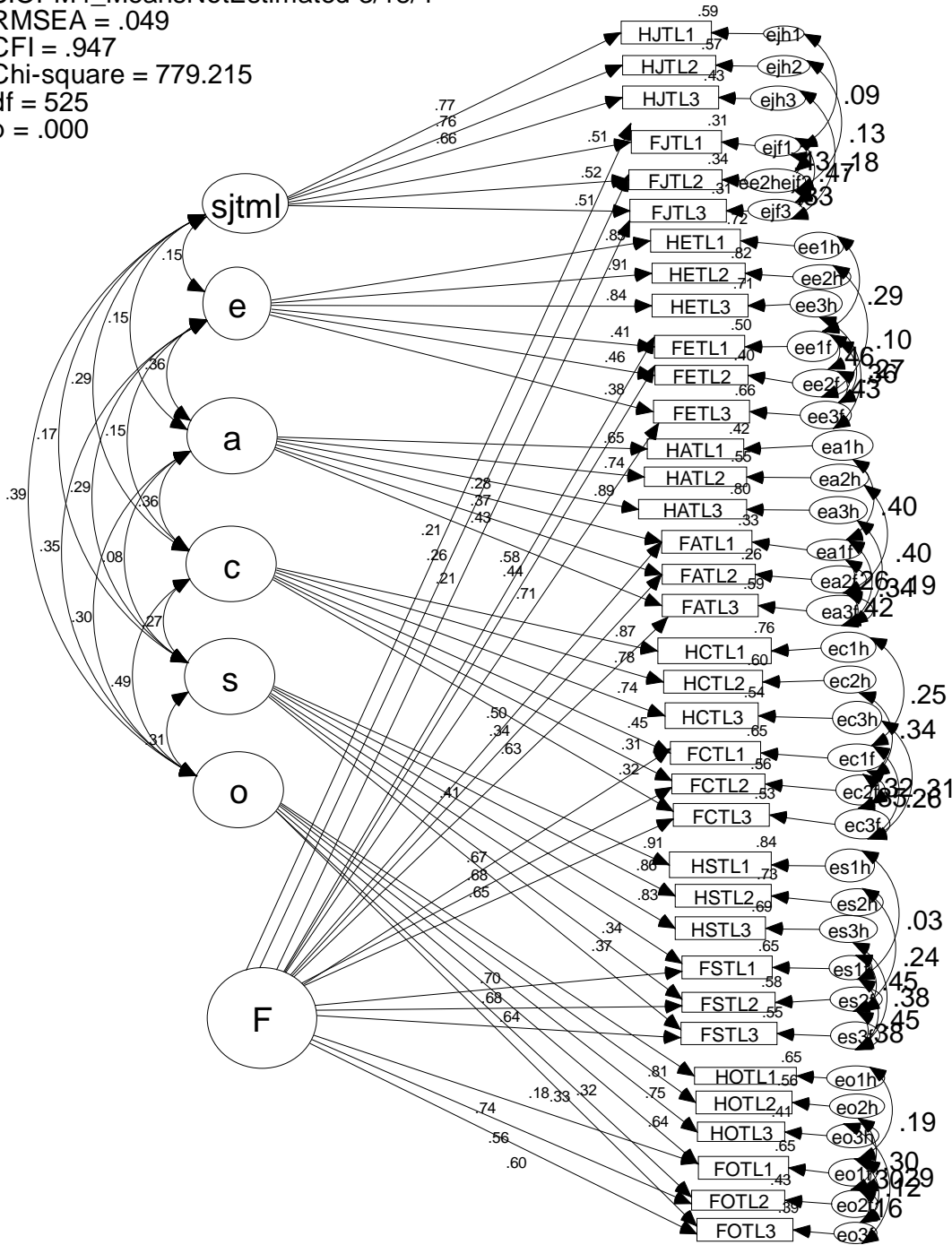


Figure 5. MTMM model applied to the data of Biderman and Nguyen (2004). In this idealized representation, each rectangle represents three testlets. Each “regression arrow” represents three such arrows from the latent variable to the three testlets represented by the rectangle. For each dimension covariances between faking condition testlet residuals were estimated as were covariances between corresponding residuals from honest and faking condition testlets. Values are means of standardized loadings or means of correlations between corresponding triplets.

MTMM Model

$\chi^2(506) = 681.8, p < .001$

CFI = .963

RMSEA = .041

Model without M

$\Delta\chi^2(19) = 97.4, p < .001$

CFI = .947

RMSEA = .049

Model without paired H~F residuals

$\Delta\chi^2(18) = 222.3, p < .001$

CFI = .920

RMSEA = .060

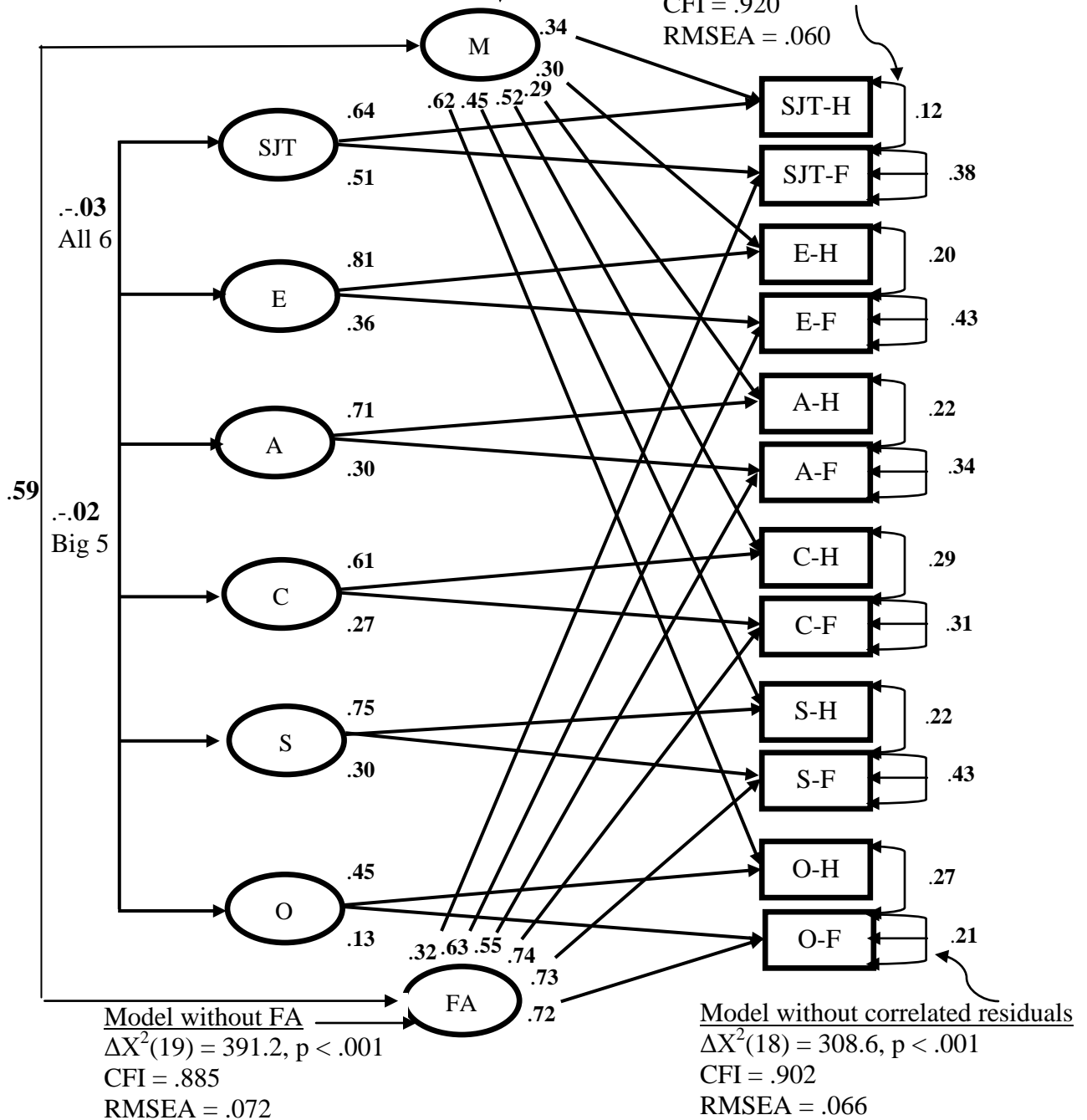


Figure 6. MTMM model applied to the data of Wrensen and Biderman (2005). In this idealized representation, each rectangle represents three testlets. Each “regression arrow” represents three such arrows from the latent variable to the three testlets represented by the rectangle. Covariances between faking condition testlet residuals were estimated as were covariances between corresponding residuals from honest and faking condition testlets. Values are means of standardized loadings or means of correlations between corresponding triplets.

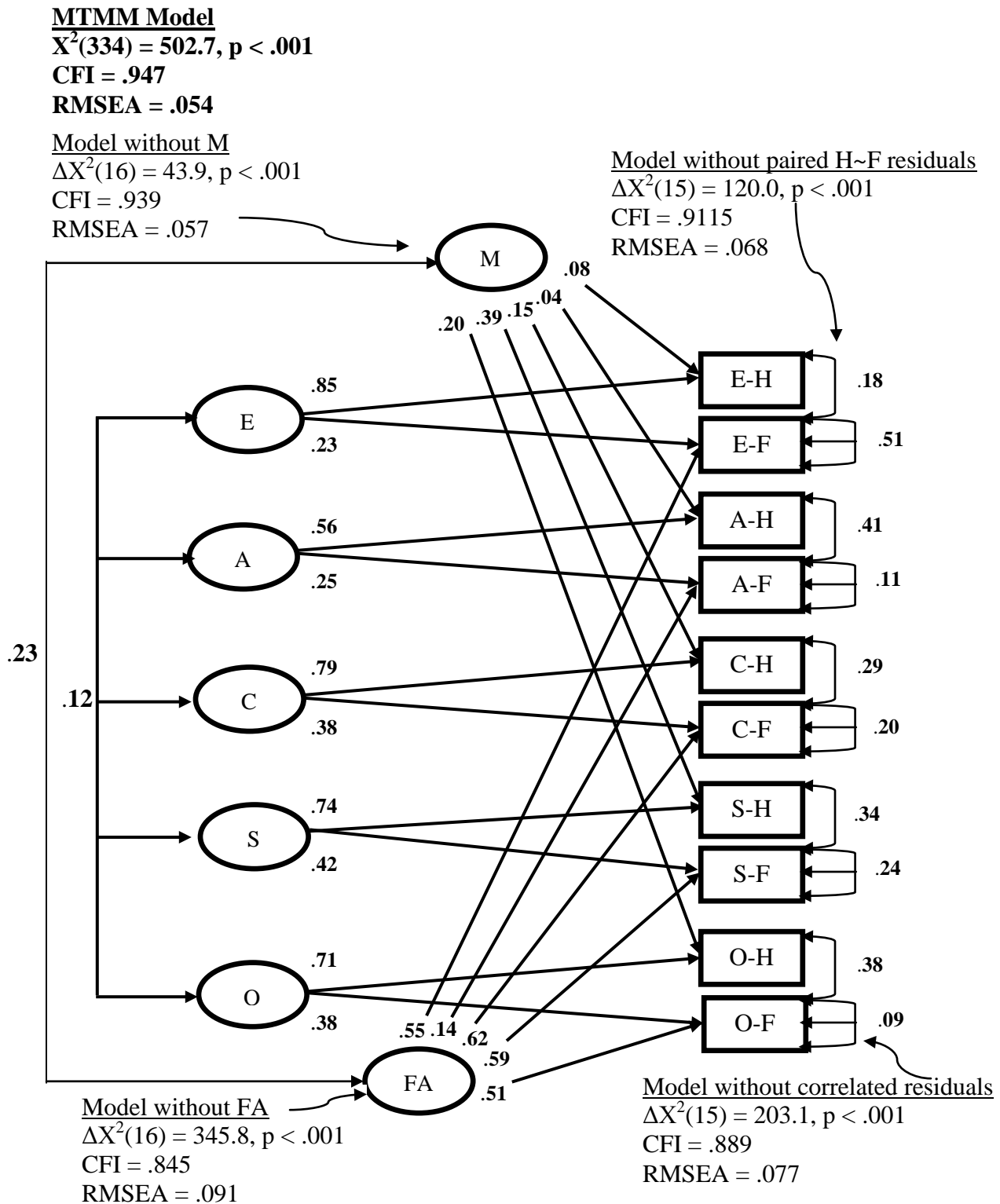


Figure 8. MTMM model applied to the Big 5 data of Clark and Biderman (2006). In this idealized representation, each rectangle represents three testlets. Each “regression arrow” represents three such arrows from the latent variable to the three testlets represented by the rectangle. D testlets were obtained in the incentive (“Dollar”) condition. I testlets were obtained in the instructed faking condition. Covariances between testlets in the D and in the I condition were estimated. Values are means of standardized loadings or means of correlations between corresponding triplets.

MTMM Model

$\chi^2(857) = 1227.6, p < .001$

CFI = .861

RMSEA = .051

Model without correlated D residuals

$\Delta\chi^2(15) = 9.4, p > .05$

CFI = .863

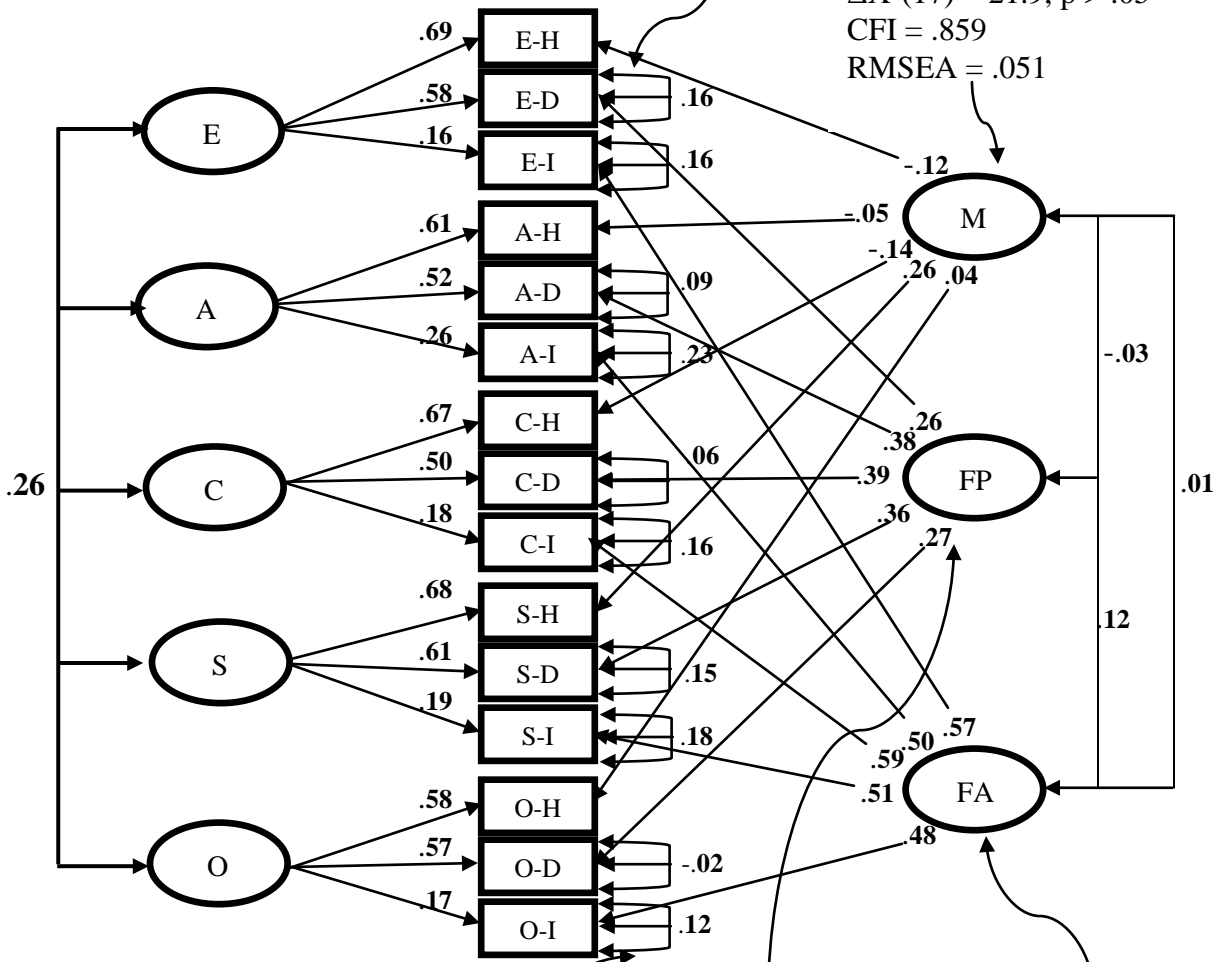
RMSEA = .050

Model without M

$\Delta\chi^2(17) = 21.9, p > .05$

CFI = .859

RMSEA = .051



Model without correlated I residuals

$\Delta\chi^2(15) = 77.7, p < .001$

CFI = .837

RMSEA = .055

Model without FP

$\Delta\chi^2(17) = 101.2, p < .001$

CFI = .829

RMSEA = .056

Model without FA

$\Delta\chi^2(17) = 328.3, p < .001$

CFI = .744

RMSEA = .068

Figure 9. Model applied to the faking condition data of the Nguyen and Biderman (2004) data.

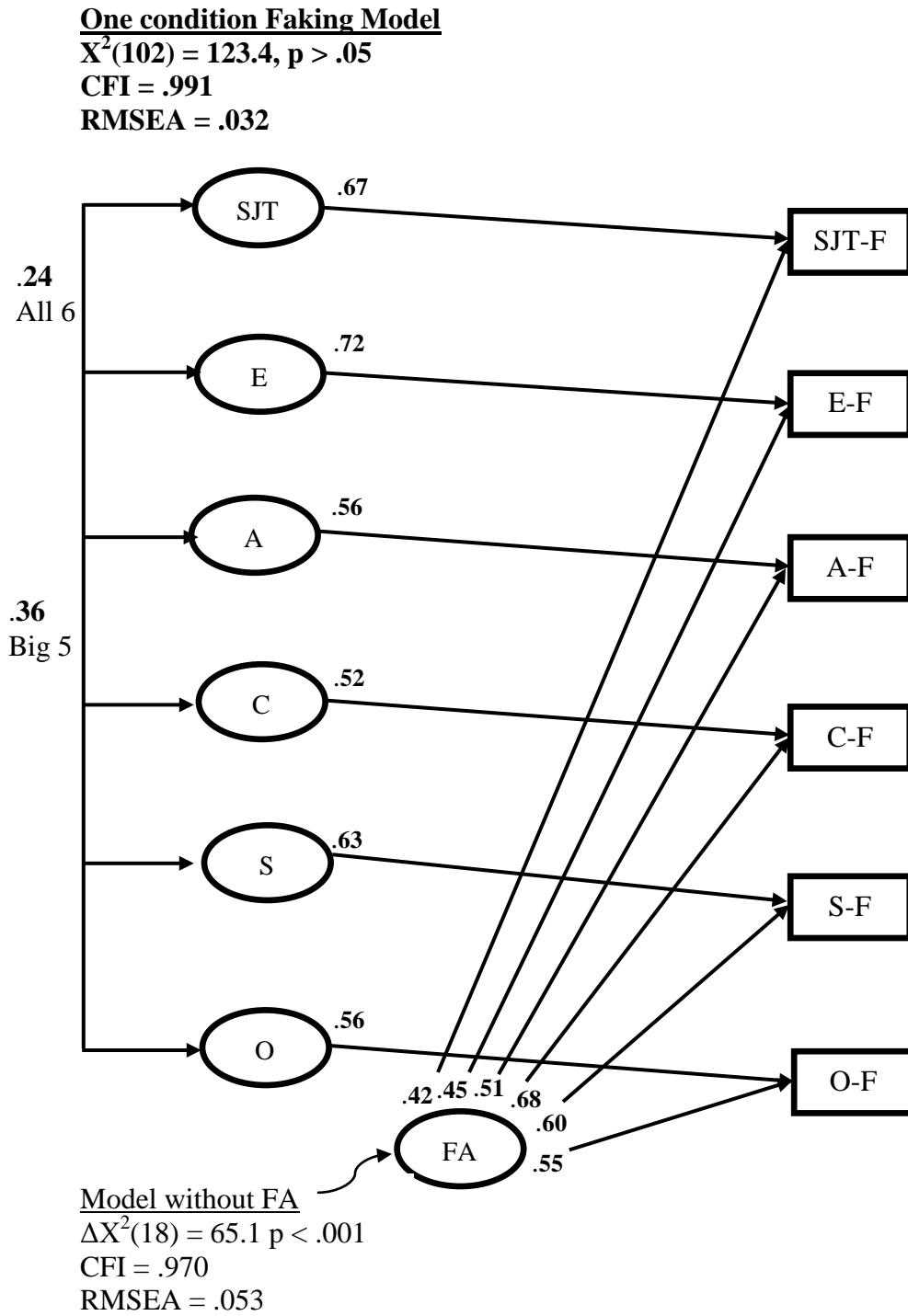


Figure 10. Application of the model to the faking condition data of the Wrensen and Biderman (2005) data.

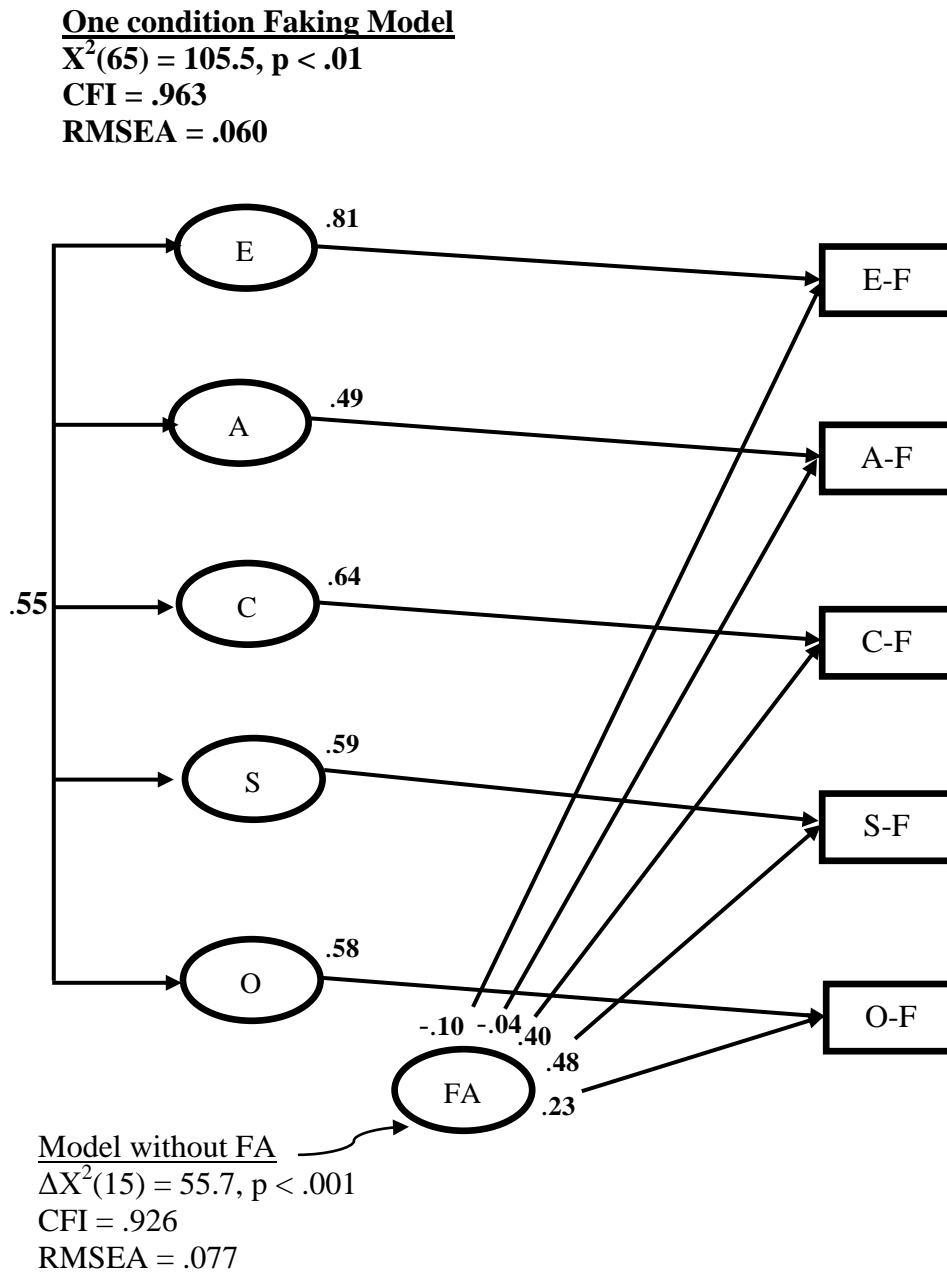


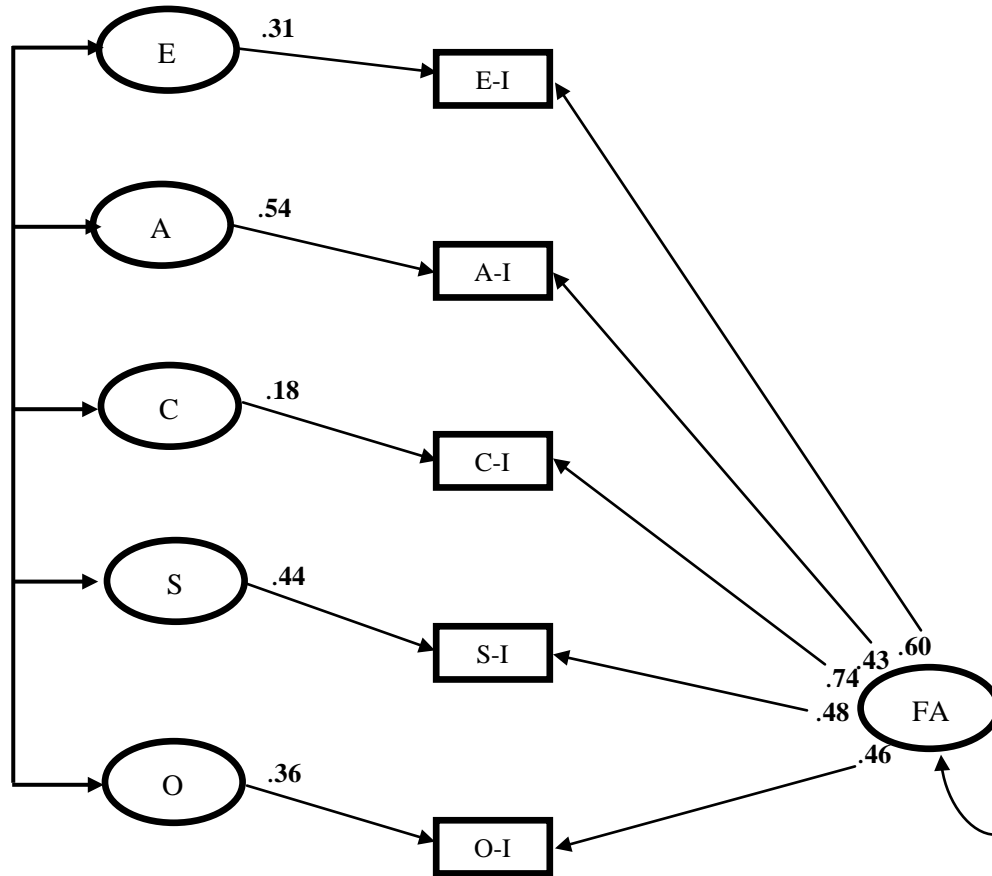
Figure 11. Application of the model to the instructed faking condition data of the Clark and Biderman (2006) data.

One condition Faking Ability Model

$\chi^2(65) = 102.8, p < .01$

CFI = .953

RMSEA = .059



Model without FA

$\Delta\chi^2(15) = 67.8, p < .001$

CFI = .886

RMSEA = .082

Figure 12. Application of the model to the incentive condition data of the Clark and Biderman (2006) data.

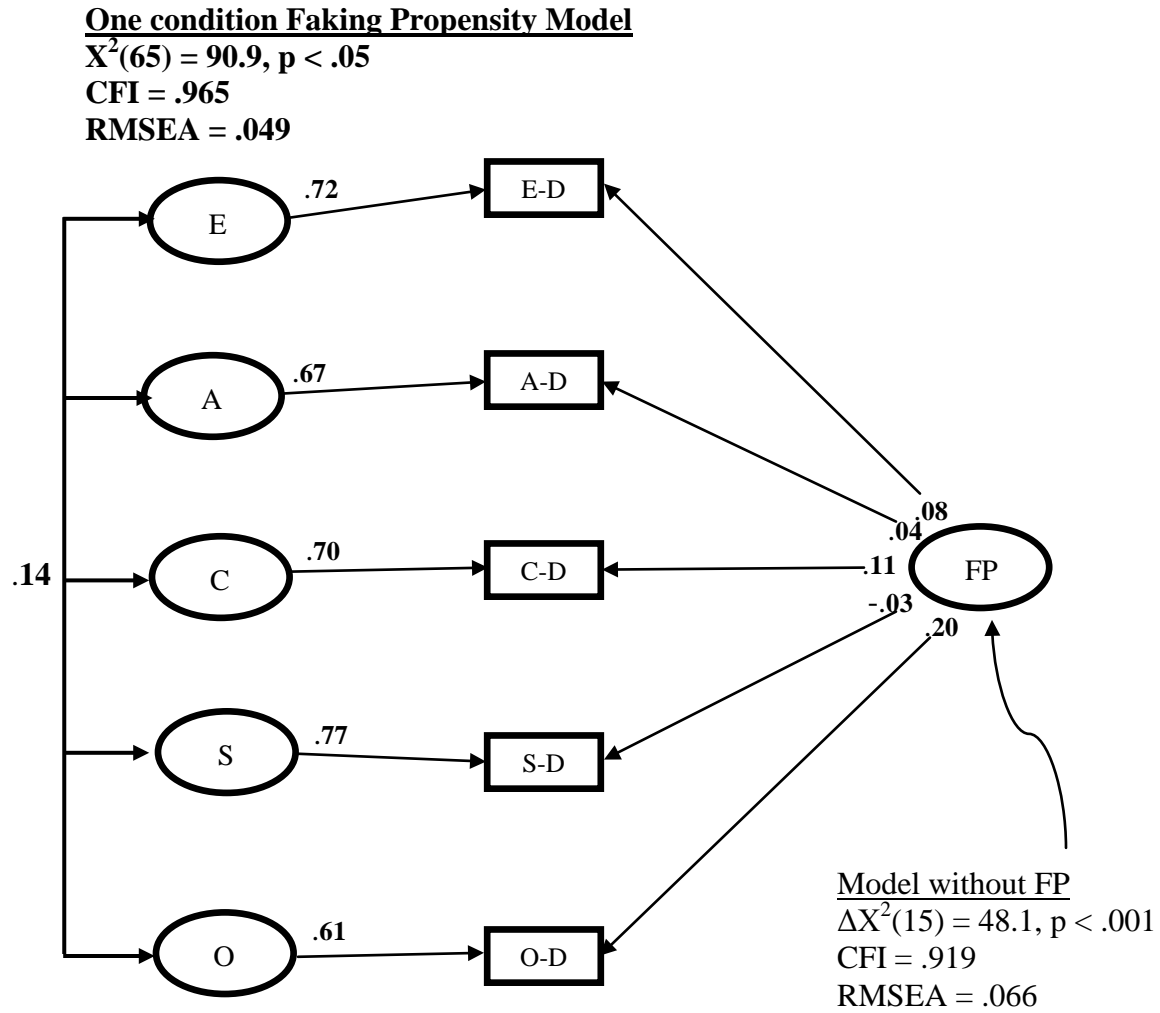
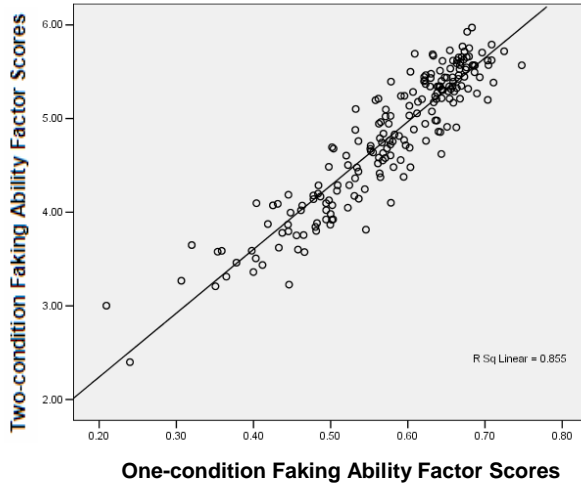
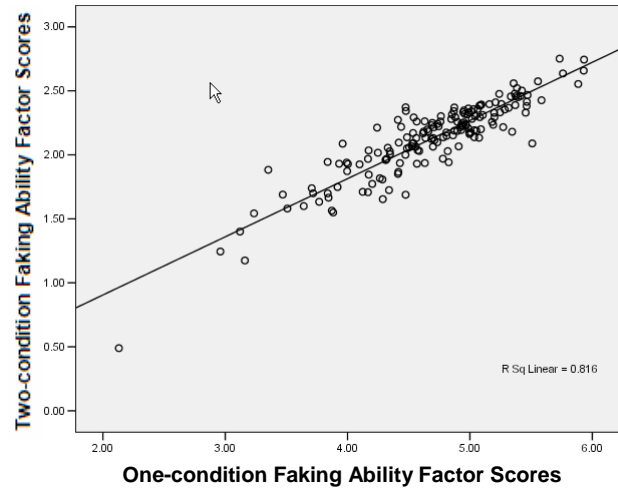


Figure 13. Scatterplots of two- and three-condition faking factor scores vs. one-condition factor scores. Panels a through c present faking ability factor scores from Biderman and Nguyen (2004), Wrensen and Biderman (2005), and Clark and Biderman (2006) respectively. Panel d presents faking propensity factor scores from the data of Clark and Biderman (2006).

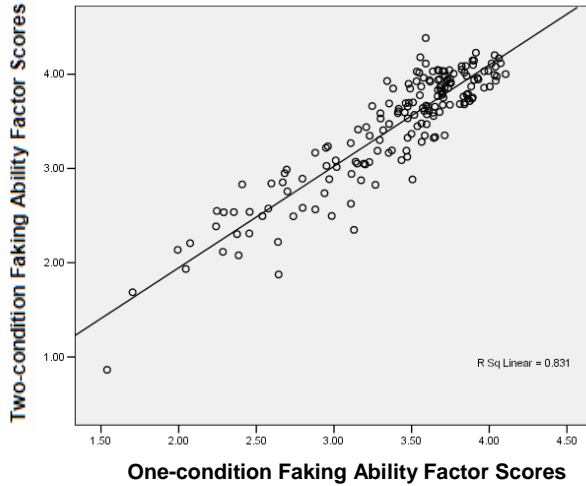
a. Biderman and Nguyen (2004)



c. Clark and Biderman (2006)



b. Wrensen and Biderman (2005)



d. Clark and Biderman (2006)

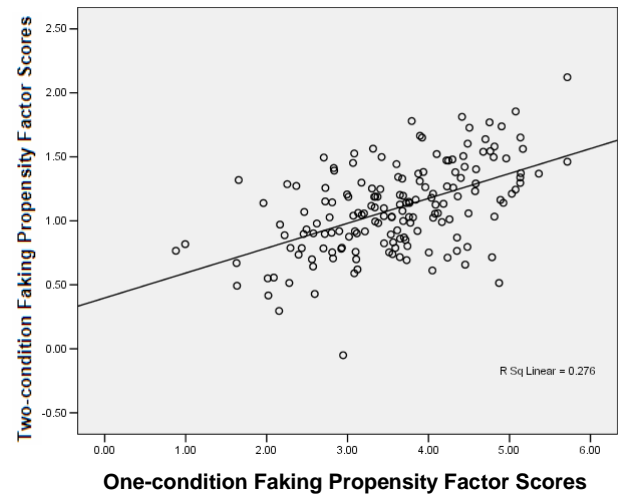


Figure 14. Amos path diagram of Biderman and Nguyen (2004) model conceptualized as a higher order multiple factor longitudinal growth model.

