Variability Indicators in Structural Equation Models

Michael D. Biderman

University of Tennessee at Chattanooga

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

Part of symposium: R. L. Griffith, Chair. T. Malm, Co-Chair. *Examining old problems with new tools: Statistically modeling applicant faking*. Conducted at the 22nd annual conference of The Society for Industrial and Organizational Psychology, New York: NY. 2007.

ABSTRACT

A latent variable indicated only by scale standard deviations was found to improve fit of structural equation models of faking of the Big Five personality dimensions. The variability latent variable was not related to faking propensity or ability. Persons with higher cognitive ability exhibited less variability within dimensions across faking conditions.

An overwhelming number of measures important in psychology are based on techniques attributed to Rensis Likert (1932 as cited in Spector, 1992). Such measures typically consist of a collection of statements of attitudes, opinions, or other personal characteristics Respondents are instructed to indicate the extent to which they agree or endorse the statements or the extent to which the statements are representative of the respondent. Presumably, a respondent reads each item and arrives at a conclusion concerning the extent of agreement with the item. This internal continuum of amount of agreement is mapped to an external response scale consisting of the first few successive integers and/or labels at equally spaced intervals. A reasonable assumption is that the respondent chooses the number or label on the external scale whose value is closest to the internal value. If several statements from the same content domain are presented, such as several statements from a single personality dimension, it seems reasonable to assume that each statement would elicit a different amount of internal agreement that would map to slightly different external responses, resulting in some variability in responses to statements from the same domain. To summarize the respondent's position on whatever dimension the statements represent, the mean or sum of external responses is obtained. That summary is used as the respondent's value or score on the dimension or variable represented by the collection of statements. It should be noted that such a summary neglects consideration of the differences between the responses to statements within the same domain – i.e., the variability of the responses mentioned above.

Most studies employing Likert-type questionnaires have summarized the responses to multiple statements or items from the same dimension in the fashion described above – by summing the responses or taking the mean of the responses. In fact, such measures are often referred to as summated rating scales (e.g., Spector, 1992). As suggested above, the focus on the central tendency of responses within a domain has been at the expense of consideration of the variability of those responses. There are relatively few studies in which the variability of behavior has been the central focus. Probably the most extensive literature is that on the concept of metatraits. A metatrait is the "quality of possessing versus not possessing a particular trait" (Britt, 1993). A related concept is traitedness, which refers to the strength of the internal representation of a trait. Traited individuals are assumed to have strong internal representations of a trait while untraited individuals have weak internal representations. (Britt, 1993; Dwight, Wolf, & Golden, 2002). Most of the studies of metatraits have used interitem variability, typically the standard deviation of responses to items within a scale, to define traitedness, with more highly traited individuals exhibiting lower interitem variability (e.g., Britt, 1993; Dwight, et. al., 2002; Hershberger, Plomin, & Pedersen, 1995). The main thrust of research on metatraits has been on traitedness as a moderator of relationships. For example, Britt (1993) found that correlations between two constructs were significantly larger for participants traited on one or both of two constructs than for those who were untraited on either or both of the constructs. But others, e.g., Chamberlain, Haaga, Thorndike, & Ahrens (2004) have found no evidence that traitedness serves as a moderator.

Another line of studies focusing on variability has examined variability of responses as extreme response style (e.g., Greenleaf, 1992). Finally, others have focused on variability of test scores across time (e.g., Eid & Diener, 1999; Kernis, 2005). Of particular interest is the study of Eid & Diener (1999) who used a confirmatory factor analysis with standard deviations of responses to the same items across days as indicators of a latent variable. This latent variable measured interindividual differences in variability across the 7 weeks of their study. A related model is proposed here for interindividual differences in variability of responses to scale items.

The paucity of uses of measures of variability is also found in the literature on factor analysis and structural equation models. Typically measurement models for latent variables involve use of scale scores, parcel scores, or item scores as indicators of the latent variables. It appears that the Eid & Diener (1999) study is the first to consider latent variables with standard deviations as indicators.

Typically, data using summated response scales are analyzed assuming that perceived agreement with a statement is the sole determinant of the responses to the statement. However, the use of personality tests for selection in industrial settings has given psychologists reason to question that assumption and to consider the possibility that factors other than only agreement might enter into the choice of response to personality items. The most frequently studied of such factors in recent years is the tendency to distort or fake responses under incentives or instructions to do so. In such situations, a reasonable assumption is that respondents add an amount to internal agreement with the response when faking or equivalently add an amount to the external response to which the internal agreement would have mapped. The result is a change in central tendency of the item under faking-inducement conditions from what central tendency of responses would have been in without the inducement to fake (e.g., McFarland & Ryan, 2000; Schmitt & Oswald, 2006). Again, this process ignores consideration of variability of responses across statements from the same domain.

The purpose of the present paper is to examine variability of responding within the context of structural equation models of faking. The impetus for the examination came from inspection of participant response sheets during data entry. It was apparent that some participants were responding in a fashion that might be best described as targeting specific responses. Reflection on this behavior lead to the hypothesis that the variability of responses of these participants would very likely be smaller than the variability of responses of persons not engaged in such targeting behavior. It also lead to the realization that this kind of behavior was not what was assumed to occur when participants filled out questionnaires under conditions conducive to faking. From this, the hypothesis that faking might be reflected as much by variability as it is by central tendency was developed.

The data to which the models investigated here were applied involved questionnaires assessing the Big Five personality dimensions (Goldberg, 1999). The questionnaires were administered in multiple-condition research paradigms consisting of a condition in which participants were instructed to respond honestly and of one or more other conditions in which instructions or incentives to fake were given. The data to which the models considered here were applied have been presented previously (Biderman & Nguyen, 2004;Clark & Biderman, 2006; Wrensen & Biderman, 2005) although the notion of issues of variability was only alluded to in one of those papers (Clark & Biderman, 2006). Clark and Biderman noted that certain aspects of the data seemed to suggest that respondents were targeting specific responses. However, models of such behavior had not yet been developed. This paper presents such models and explores how they might account for such targeting behavior.

The core structural equation model for the two-condition faking paradigm originally presented by Biderman and Nguyen (2004) is presented in Figure 1. The figure shows the model applied to the data of the Big Five personality inventory, the questionnaire used in all the datasets presented here. The two-condition faking paradigm shown in the figure is one in which participants are given the same questionnaire or two equivalent questionnaires in two experimental conditions – once with instructions to respond honestly and then again with incentives or instructions to distort their responses. Biderman and Nguyen (2004) proposed that faking be modeled by adding a latent variable representing individual differences in the amount participants added to each item in the faking condition. This latent variable is denoted F in the figure.

Biderman and Nguyen (2004) found that adding the F latent variable to the model significantly increased goodness-of-fit. This result was consistent with the hypothesis that there are individual differences in amount of distortion by respondents in the instructed faking situation. Biderman and Nguyen found that these differences were positively related to cognitive ability as assessed by the Wonderlic Personnel Test (WPT: Wonderlic, 1999). In a second application of the model using an instructed faking paradigm Wrensen and Biderman (2005) found that faking ability was again positively related to cognitive ability and also positively related to scores on measures of emotional intelligence and integrity and negatively related to a measure of socially desirable responding. Finally, Clark and Biderman (2006) applied the model to whole scale scores of a within-subjects paradigm involving an honest-response condition, a condition with incentive to fake, and a condition with instructions to fake. In this application, two faking latent variables were estimated. The first, called FP for faking propensity, represented individual differences in distortion in the incentive condition. The second, called FA for faking ability, represented individual differences in distortion in the instructed faking condition. The model applied by Clark and Biderman (2006) is presented in Figure 2. As was found in the first two studies, addition of the faking latent variables significantly improved goodness-of-fit. The data of this study are consistent with the hypothesis that there are individual differences in the propensity to fake and also individual differences in the ability to fake. Interestingly, the correlation between the two faking latent variables was not significantly different from zero, a finding that certainly merits further research. Taken together, the results of these three applications of the faking model suggest that faking or response distortion can be at least partially represented by the additive model originally presented by Biderman and Nguyen (2004). However, the question of whether or not respondents to questionnaires engage in other forms of distortion such as targeting remains.

METHOD

Datasets.

The data of three different samples involving administration of a Big Five questionnaire were analyzed. The first dataset was that reported upon by Biderman & Nguyen (2004; see also Nguyen, 2002; Nguyen, Biderman, & McDaniel, 2005). It was comprised of 203 undergraduate and graduate student participants from two southeastern universities. Participants were given a Situational Judgment Test and the Big 5 questionnaire twice, once with instructions to respond honestly and again with instructions to respond in a fashion that would increase the participant’s chances of obtaining a customer service job. Half the participants were given the honest condition first. Only the Big Five data of this sample were analyzed here. Participants were given the WPT (Wonderlic, 1999)) prior to the experimental manipulation.

The second dataset was similar to the first, with an honest-response and a fake-good condition (Wrensen & Biderman, 2005) with order of presentation of the conditions counterbalanced. Several other questionnaires including the WPT were given prior to the experimental manipulation. Sample size was 166.

In the third dataset (Clark & Biderman, 2006), participants were exposed to three conditions. In the first, they were instructed to respond honestly. In the second, incentive condition (also referred to as the Dollar condition in figures), they were told that the names of persons whose scores were most likely to be appropriate for a customer service job would be entered into a prize drawing for a $50 gift certificate. In the third condition, the same participants were instructed to respond in a fashion that would be most likely to gain them employment in a customer service job. In this study order of exposure to the conditions was the same for all participants – honest followed by incentive followed by instructed faking. Sample size was 166.

Measures

Goldberg's Big Five Personality Inventory (Goldberg, et. al., 2006). For the Biderman and Nguyen (2004) and Wrensen and Biderman (2005) studies, the Goldberg 50-item Big Five questionnaire available on the IPIP web site was used. Ten items from the questionnaire mark each of the dimensions, Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Imagination/Intellect. Participants responded indicating how accurate each item was as a description of themselves on a five-point scale, from “Very inaccurate” to “Very accurate”. For the Clark and Biderman (2006) study, items from the 100-item scale available from the same web site were divided into three approximately equivalent 30-item forms. The order of administration of these forms was counterbalanced across the three conditions of this study – Honest, Incentive, and Instructed faking. The forms were treated as equivalent for the purpose of the analyses reported here. Participants responded by indicating how accurate each item was as a description of themselves on a seven point scale. Some of the items were slightly reworded to decrease the likelihood of ceiling effects in the faking conditions. . These items were reworded to make it less likely that they would be rated as “Very accurate” descriptions if they were positively worded or “Very inaccurate” if they were negatively worded. The modifications consisted of adding only adjectives such as, “always,” “never,” and “sometimes” to the statements.

The items of the first two datasets were grouped into five two-item parcels per dimension. Each parcel consisted of the mean of the 1st and 6th, 2nd and 7th, 3rd and 8th and so on for the 10 items representing each dimension using the order of items as presented on the IPIP web site.[1] Although there is some disagreement in the literature concerning the appropriateness of using parcels, our take on the literature is that parceling is appropriate as needed when items are unidimensional (Sass & Smith, 2006). Since the Big Five items used for this study have a long history of development it seems appropriate to assume that the items within each scale are unidimensional. For the third dataset, whole scale scores consisting of the means of six items per dimension were analyzed.

Targeting. To address the issue of targeted responses, it was decided that targeting would best be represented by low variability of responses to items within a Big Five dimension. To measure such variability, in keeping with the past research involving metatraits, standard deviations of items within the Big Five dimension were computed. Each standard deviation was computed from all 10 responses to the items in a dimension in the first two studies and from all six responses in the third. A standard deviation was computed for each instructional condition. This meant that for each participant in the first two studies, two standard deviations were computed for each Big Five dimension - one for the honest instructional condition and one for the instructed faking condition. For the third study, three were computed for each dimension for each participant - one for the honest, incentive, and instructed faking conditions respectively. These standard deviations were added to the data to be modeled.

Wonderlic Personnel Test (WPT). All participants took Form A of the Wonderlic personnel test. The WPT was included as an exogenous variable in structural model presented later.

Model.

The model applied was a generalization of the Faking model originally presented by Biderman and Nguyen (2004). The parcel or whole scale means were modeled in the fashion described above and illustrated in Figures 1 and 2[2].

The standard deviations were linked to the other variables in the models in two ways. First, each standard deviation was regressed onto its corresponding central tendency indicator(s), i.e., scale score or set of parcels. For example, for a dataset for which whole scale scores were analyzed, the standard deviation of the extroversion items was regressed onto the extroversion score, the standard deviation of agreeableness items onto the agreeableness score, and so forth. For datasets for which parcels were analyzed, the standard deviation for each dimension was regressed onto the set of parcels representing that dimension. These *regression links* were designed to account for relationships that might occur when a shift in central tendency moved the distribution of responses for an individual near the end of the external response scale leading to a ceiling effect that would reduce variability. Since the faking conditions employed here involved incentives or instructions to fake good, it was expected that the relationships found here would be negative, with increases in central tendency due to faking associated with decreases in variability.

The second way in which the standard deviations were linked with other variables was through the introduction of a latent variable (V) for which the standard deviations were the indicators. The V latent variable was designed to represent individual differences in variability of responses. Since the standard deviations were also regressed onto scale central tendency through the regression links, differences in variability that might have been associated with ceiling effects were partialled out. Standard deviations from all conditions served as indicators of V. V was allowed to correlate with faking latent variables and to the Big Five latent variables. Figures 3 and 4 present this *Variability* model as applied to the first two datasets and to the 3rd dataset respectively[3].

The models were applied using Mplus Version 4.2 (Muthen & Muthen, 1998-2006) and Amos Version 6 (Arbuckle1995-2005). A sequence of models was applied, and differences between goodness-of-fit of the models was tested using chi-square difference tests. First a model including both regression links and the V latent variable was applied. This model was labeled Model 3. Then a model without the V latent variable, Model 2, was applied. Next Model 1 with the V latent variable but without regression links was applied. Finally, in Model 0, neither regression links nor the V latent variable was included, leaving the standard deviations unmodeled. Each restricted model could be formed from Model 3 by restricting a set of parameters to 0. Thus the chi-square difference test is appropriate to compare the fit of each restricted model to the more general Model 3.

RESULTS

Path diagrams of the application of Model 3 to the data of each sample are presented in Figures 7 through 9. In the figures, means of loadings or regression coefficients are presented for each dimension. Inspection of the figures suggests that the standard deviations were negatively related to central tendency and that standard deviations loaded positively on the V latent variable.

Table 1 presents goodness-of-fit measures for the four models applied to each dataset along with chi-square difference tests comparing the fit of each restricted model to Model 3.

Eliminating the regression links in Model 1 resulted in significant decrements in goodness-of-fit for all three samples. This indicates that there were relationships between variability and central tendency in all three samples. The generally negative regression coefficients are consistent with ceiling effects.

The comparison of Model 2 vs. Model 3 tested the viability of the V latent variable. As Table 1 shows, eliminating the V latent variable in Model 2 resulted in significant decrements in goodness-of-fit. This result is consistent with the positive loadings of all the standard deviations on V.

Finally, Model 0 left the standard deviations unmodeled. As might be expected, it also fit significantly worse than Model 3, suggesting the need for both regression links and the V latent variable.

Table 2 presents model information on the relationship of variability to faking. The table presents loadings of the V parameter on the honest condition and faking condition indicators. If participants were targeting responses in the conditions involving incentives or instructions to fake, we would expect individual differences in participant variability of responding to have greater impact in the faking conditions than in the honest conditions. This leads to the expectation that if participants were targeting responses in the faking conditions, loadings of indicators from those conditions would be larger than loadings from the honest conditions. To simplify the table, means of standardized loadings across parcels within dimension are presented for datasets 1 and 2. Inspection of the table suggests that the loadings are essentially equivalent in honest and faking conditions.

If variability of responding were associated with faking, we would expect a negative relationship between the V latent variable and the Faking latent variables, with those most able or inclined to alter their responses exhibiting smaller variability of responding than those participants who exhibited less response distortion. Correlations of the V latent variable with the Faking latent variables were .04 and .02 ($p > .05$ for both) in the first two studies. In the third study, the correlation of V with FP was .16 and with FA was .10 ($p > .05$ for both). Thus, while there is evidence that different participants exhibited different amounts of variability, these results lead to the tentative conclusion that those differences were not related to response distortion under incentives or instructions to fake. Instead, the evidence suggests that individual differences in variability of responding were essentially constant across instructional conditions.

There is a saying attributed to Mark Twain that says “To a man with a hammer, everything looks like a nail.” The consistency of individual differences in variability of responding in the three datasets can certainly be considered a hammer in search of a nail. Since the three datasets were not gathered to study differences in consistency, there are certainly variables that should have been included in them that were not. The only measure that was taken across all three datasets was the WPT. It was administered to investigate the relationship of response distortion to cognitive ability. In the original reports of these datasets it was found that WPT scores were positively related to faking ability in all three studies, although only weakly so in the third. Since the WPT scores were available, an exploratory study of the relationship of variability to cognitive ability was conducted. For this examination, WPT scores were included in the model and allowed to covary with all of the latent variables (E, A, C, S, I, the faking latent variables, and V) in the model. The correlations, critical ratios and p-values are presented in Table 3. In all three datasets, the correlation of WPT scores with V was negative, with $p < .05$ in two studies and $p < .10$ in the Wrensen & Biderman (2005) study. In each dataset, persons with higher cognitive ability exhibited lower variability of responses across items within dimensions.

DISCUSSION

This research examined an alternative dimension of responding to personality tests. Much previous research has focused on the central tendency of responses to personality test items – on scale mean scores. This research examined instead the variability of responses to the items of each personality dimension and found that the variability may provide a useful perspective that is not afforded by scale means alone.

The first major finding in this research was that variability of responding is a consistent individual difference variable. The V latent variable increased the goodness-of-fit of the model considerably in three separate datasets when compared to a model with only regression links between scale means and scale standard deviations. This means that some individuals consistently gave nearly identical responses to all items within each scale while others consistently distributed their responses more widely. The generally significant loadings of all the standard deviation indicators on V suggested that this tendency cut across scales and across faking conditions.

The second major result is the tentative conclusion that variability of responding is not related to faking. Although none of the results present above suggested that targeting behavior occurred, because this is the first application of the variability model we are hesitant to reach a definitive conclusion concerning how respondents behave when faking. There are many variations on the model presented here that could result in different conclusions. For example, we have not yet had time to thoroughly explore models involving more than one Variability latent variable, one for each condition, for example. Along with that possibility is the possibility that mean variability may change from one condition to the next. Both of these alternative models have yet to be examined. For that reason, we regard the conclusion that variability of responding is unrelated to faking as tentative.

The third major finding is the fairly consistent relationship of variability to cognitive ability. Participants higher in cognitive ability exhibited less variability in responding to items within scales. We note that cognitive ability was also related to faking ability. Table 3 shows that this relationship was significantly positive in two of the studies and marginally so in the third. Table 3 also shows that cognitive ability was related to Imagination/Intellect in all three studies, significantly so in two and marginally in the third. This general pattern of results suggests that cognitive ability is clearly related to behavior of respondents to personality questionnaires. To explore the extent to which cognitive ability is related to response characteristics from the Big Five, a structural model was created in which the WPT scores were regressed onto the V, F, and O latent variables. For the three datasets, the multiple r's were .57, .35, and .50 respectively. Although these correlations are not large enough to suggest that instruments specifically designed to measure cognitive ability should be replaced with specially scored personality tests, they do indicate that there is potentially useful information about cognitive ability in a popular Big Five questionnaire.

Although variability of responding appears to be related to cognitive ability, it is not consistently related to the Big Five dimensions. None of the correlations of V with E, A, C, or S was significant. Only the correlation of V with I from the Wrensen & Biderman (2005) study reached the .05 level of significance. Given the relationship of both to cognitive ability, this last correlation is not surprising. This pattern of insignificant correlations suggests that the variability characteristic has discriminant validity with respect to the Big Five, although, of course, tests of its convergent validity as yet lack a theoretical foundation.

The significant improvements in model fit from the regression links introduced in Model 2 suggest that researchers investigating traitedness as a moderator variable should consider whether or not some of the differences in variability ascribed to differences in traitedness are due to floor or ceiling effects associated with extremes in central tendency. Some previous studies (e.g., Hershberger, et. al, 1995) have considered such effects. Other, e.g., Dwight, et. al., 2002) apparently have not. If items are chosen so that the mean of the distribution of responses is near the middle of the response scale, then artifactual differences in variability may not be problematic. But in research in which the distribution of responses is near the ceiling or floor of a response scale or may be moved there due to research conditions, such as those found in faking studies, it will be important to partial out the effects of changes in central tendency on variability when studying the effects of traitedness.

The discovery of an individual difference characteristic buried in data that had been analyzed previously makes one wonder what other characteristics may be hiding in the data that we, perhaps too superficially, analyze using standard summated scale techniques. For example, consider a "relatedness" construct defined as the correlation of a respondent's responses to the same items presented in two situations. No claim is made here that such a construct or others like it would be useful, but, like the variability construct presented here, such constructs represent aspects of responses to questionnaire items that may be in data we have already collected. Who knows what use we might make of them?

REFERENCES

Arbuckle, J. L. (1995-2005). *Amos 6.0 User's Guide.* Chicago: SPSS, Inc.

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.

Muthén, L.K. and Muthén, B.O. (1998-2006). *Mplus User's Guide. Fourth Edition.* Los Angeles, CA: Muthén & Muthén.

Britt, T. W. (1993). Metatraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and social Psychology, 65*, 554-562.

Chamberlain, J. M., Haaga, D. A., Thorndike, F. P., & Ahrens, A. H. (2004) Metatraits and assessment of attributional style. *Journal of Psychology: Interdisciplinary and Applied. 138,* 521-531.

Clark III, J. M., & Biderman, M. D. (2006). A structural equation model measuring faking propensity and faking ability. Paper presented at the 21st annual conference of the Society for Industrial and Organizational Psychology. Dallas, TX - May.

Dwight, S. A., Wolf, P. P., & Golden, J. H. (2002). Metatraits: Enhancing criterion-related validity through the assessment of traitedness. *Journal of Applied Social Psychology, 32*, 2202-2212.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76,* 662-676.

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

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Clonginger, 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.

Greenleaf, E. a. (1992). Measuring extreme response style. *The Public Opinion Quarterly, 56,* 328-351.

Hershberger, S. L., Plomin, R., & Pedersen, N. L. (1995). Traits and metatraits: Their reliability, stability, and shared genetic influence. *Journal of Personality and Social Psychology, 69,* 673-685.

Hiscock & Inch (1996)

Kernis, M. H. (2005). Measuring self-esteem in context: The importance of stability of self-esteem in psychological functioning. *Journal of Personality, 73,* 1569-1605.

Moreno, Buchanan, Van-Orden, & Guy (2002),

Likert, R. (1932). "A technique for the measurement of attitudes." *Archives of Psychology, 22: No. 140*.

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

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.

Nguyen, H., Biderman, M., & McDaniel, M. (2005). Effects of Response Instructions on Faking a Situational Judgment Test. *International Journal of Assessment and Selection*, *13*, 250-260.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightman (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Sass, D. A., & Smith, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling, 13*, 566-586.

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 2006*, 91, 613-621.

Spector, P. E. (1992). *Summated Rating Scale Construction: An introduction.* Sage University Paper series on quantitative Applications in the Social Sciences, series no. 07-082. Newbury Park, CA: Sage.

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. April.

FOOTNOTES

[1] Recently, the analyses reported here have been conducted using individual items as indicators. All of the key results presented here were replicated in those analyses.

[2] For the first two datasets, residuals between identical parcels from the honest and faking conditions were allowed to correlate. In addition, residuals of the faking parcels were also allowed to correlate. These additions significantly improved the fit of the model to the parcel data and were described by Biderman and Nguyen (2004).

[3] The original expectation associated with targeting behavior was that it was specific to the faking conditions. This belief was based on selective perceptions of respondent answer sheets during manual entry of data. In the original application of the model, the V latent variable was indicated only by standard deviations from the faking conditions. In what can only be described as a moment of high openness to experience, a decision to see if the honest condition standard deviations also loaded on V was made, leading to the current model in which all standard deviations were allowed to load on V.

Table 1 Goodness-of-fit of four models.. In Model 0, standard deviations were not modeled. In Model 1, standard deviations were modeled only by regression links. In Model 2 standard deviations were modeled only by the V latent variable. Model 3 standard deviations were modeled by both regression links and the V latent variable. Chi-square differences compare each model to Model 3.

Sample 1: Biderman & Nguyen (2004). N=203.[a]

Model	X^2	df	CFI	RMSEA	X^2 Diff	p<	Description
3	2501.249	1539	.871	.055			V, Regs
2	2719.290	1555	.843	.061	218.041	.001	No V
1	3126.934	1589	.793	.069	625.685	.001	No regs
0	3407.247	1605	.758	.074	905.998	.001	No V, regs

Sample 2: Wrensen & Biderman (2005). N=166.[b]

Model	X^2	df	CFI	RMSEA	X^2 Diff	p<	Description
3	2666.874	1539	.816	.066			V, Regs
2	2941.342	1555	.774	.073	274.468	.001	No V
1	3414.123	1589	.702	.083	747.249	.001	No regs
0	3717.764	1605	.655	.089	1050.890	.001	No V, regs

Sample 3: Clark & Biderman (2006). N=166.

Model	X^2	df	CFI	RMSEA	X^2 Diff	p<	Description
3	532.552	352	.883	.056			V, Regs
2	937.885	374	.633	.095	405.333	.001	No V
1	800.919	367	.718	.084	268.367	.001	No regs
0	1152.146	399	.510	.107	619.594	.001	No V, regs

Table 2. Doubly standardized loadings and critical ratios of standard deviations on the Variability latent variable across conditions. All p-values < .001.

Biderman & Nguyen (2004)

	Honest		Incentive to fake		Instructed to fake	
Dimension	Loading	CR	Loading	CR	Loading	CR
Extraversion	.309	-			.267	2.738
Agreeableness	.515	3.713			.418	3.554
Conscientiousness	.471	3.652			.345	3.484
Stability	.360	3.203			.348	3.250
Imag/Intell	.375	3.441			.451	3.618

Wrensen & Biderman (2005)

	Honest		Incentive to fake		Instructed to fake	
Dimension	Loading	CR	Loading	CR	Loading	CR
Extraversion	.521	-			.557	5.806
Agreeableness	.391	5.010			.361	5.367
Conscientiousness	.354	4.244			.402	5.609
Stability	.376	4.125			.254	3.373
Imag/Intell	.415	4.716			.463	5.729

Clark & Biderman (2006)

	Honest		Incentive to fake		Instructed to fake	
Dimension	Loading	CR	Loading	CR	Loading	CR
Extraversion	.461	-	.544	4.825	.391	4.173
Agreeableness	.435	4.219	.509	4.487	.431	4.417
Conscientiousness	.391	3.906	.405	3.984	.329	3.659
Stability	.511	4.549	.501	4.545	.491	4.530
Imag/Intell	.540	4.715	.645	5.152	.423	4.256

Table 3. Correlations and critical ratios of latent variables with the WPT.

Biderman & Nguyen (2004)

	V	FA	FP	E	A	C	S	I
r with WPT	-.24[a]	.42[c]		.01	-.08	.13	.00	.44[c]
Critical ratio	2.32	4.51		0.12	0.98	1.73	0.04	4.93

Wrensen & Biderman (2005)

	V	FA	FP	E	A	C	S	I
r with WPT	-.17	.18[a]		.05	-.04	.04	.21[a]	.17
Critical ratio	1.91	2.09		0.58	0.51	0.48	2.39	1.94

Clark & Biderman (2006)

	V	FA	FP	E	A	C	S	I
r with WPT	-.26[b]	.11	-.09	.12	-.09	.06	.12	.39[c]
Critical ratio	2.69	1.26	0.76	1.37	0.91	0.58	1.33	3.96

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

Figure 1. Basic faking model presented by Biderman and Nguyen (2004) as applied to data consisting of five parcels per Big Five dimension. The dimensions are Extraversion, Agreeableness, Conscientiousness, Stability, and Imagination/Intellect. Residual latent variables are not shown.

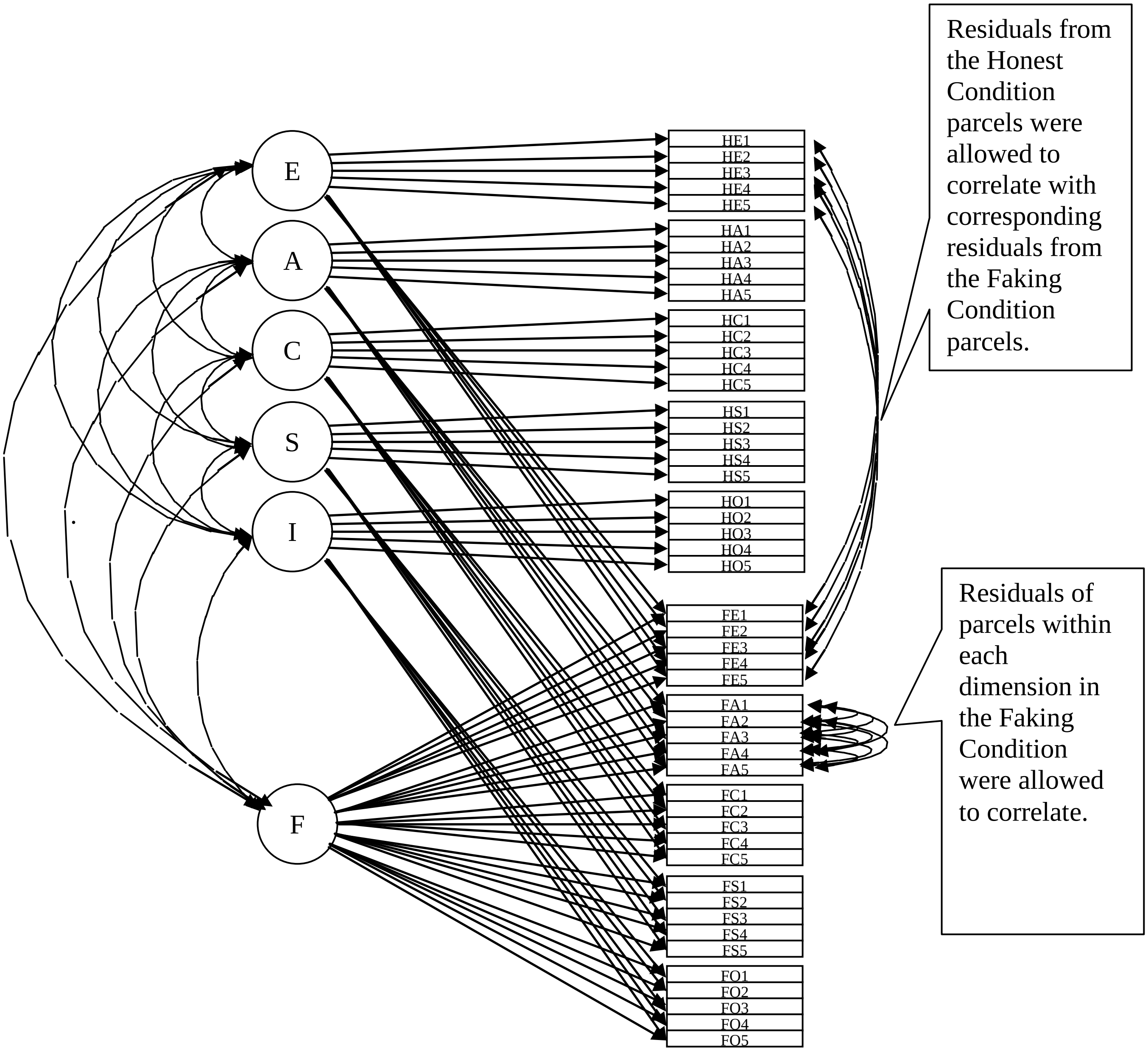


Figure 2. Clark and Biderman (2006) model applied to Big Five whole-scale scores. H represents the honest response condition, D the incentive (or Dollar) condition, and I the instructed faking condition. The dimensions are Extraversion, Agreeableness, Conscientiousness, Stability, and Imagination/Intellect. Residual latent variables are not shown.

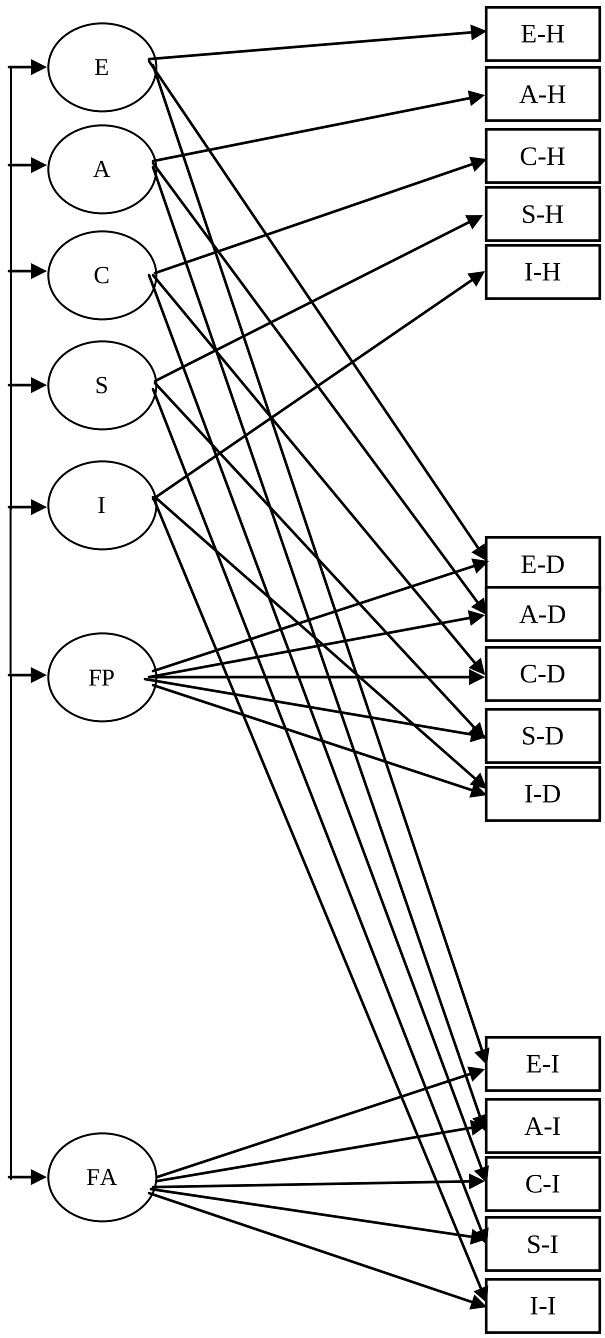


Figure 3. Basic faking model with regression links to scale standard deviations added. Residual latent variables and correlations among them are not shown.

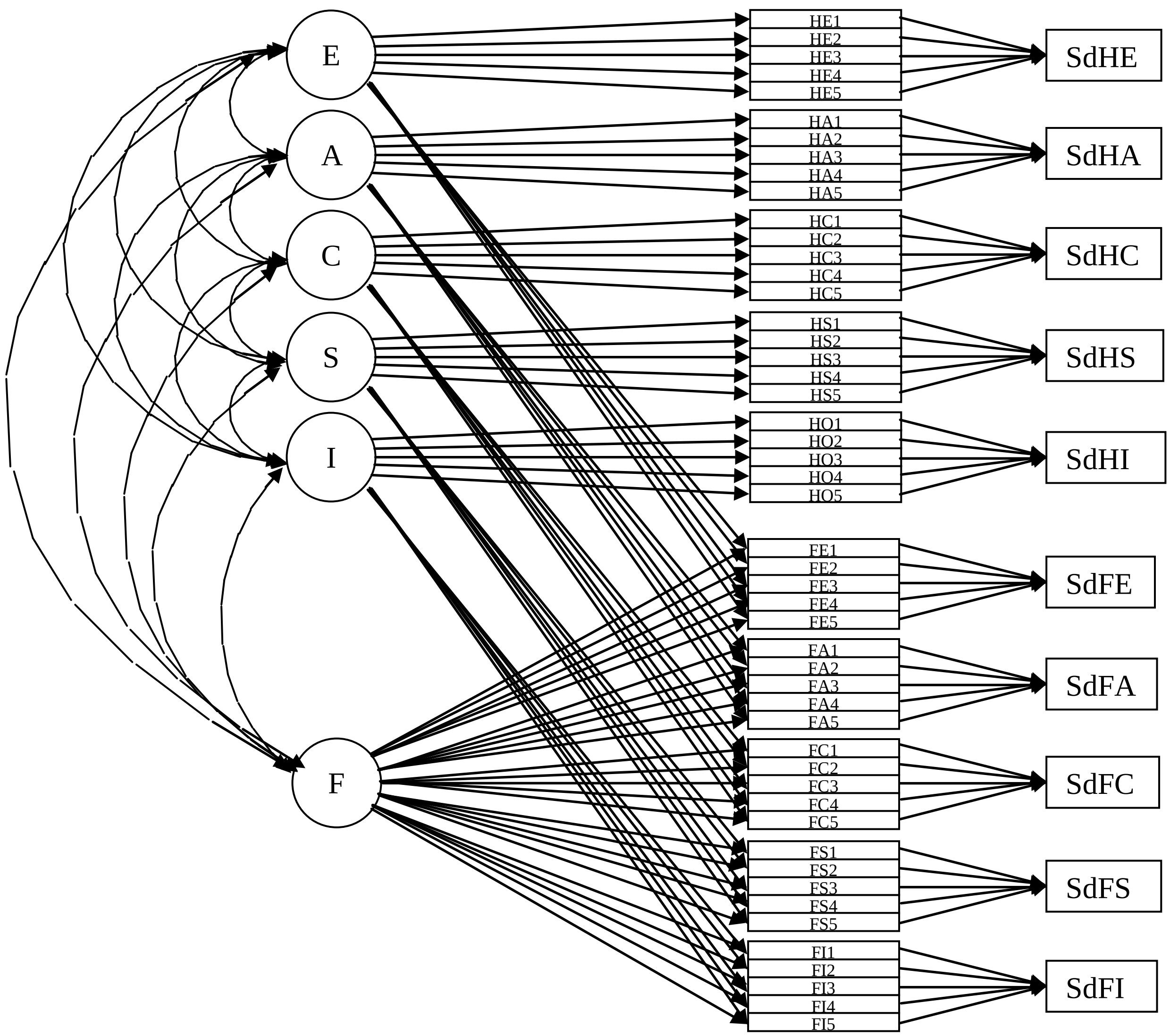


Figure 4. Clark and Biderman (2006) model with regression links to scale standard deviations added. Residual latent variables are not shown.

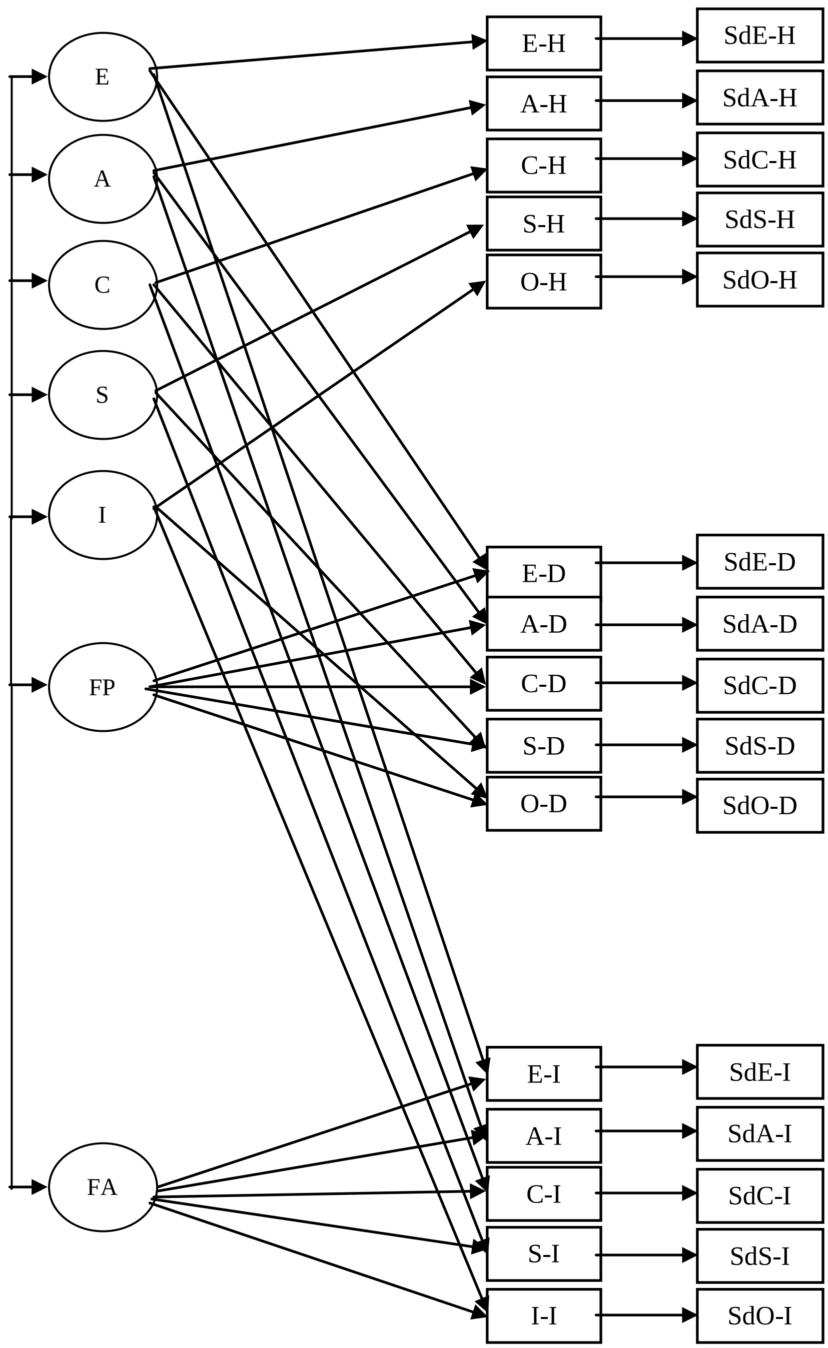


Figure 5. Complete variability model applied to five-parcel data of Biderman and Nguyen (2004) and Wrensen and Biderman (2005). Residual latent variables and correlations between them are not shown. Although not shown in the figure, V was allowed to correlate with F and the Big Five latent variables.

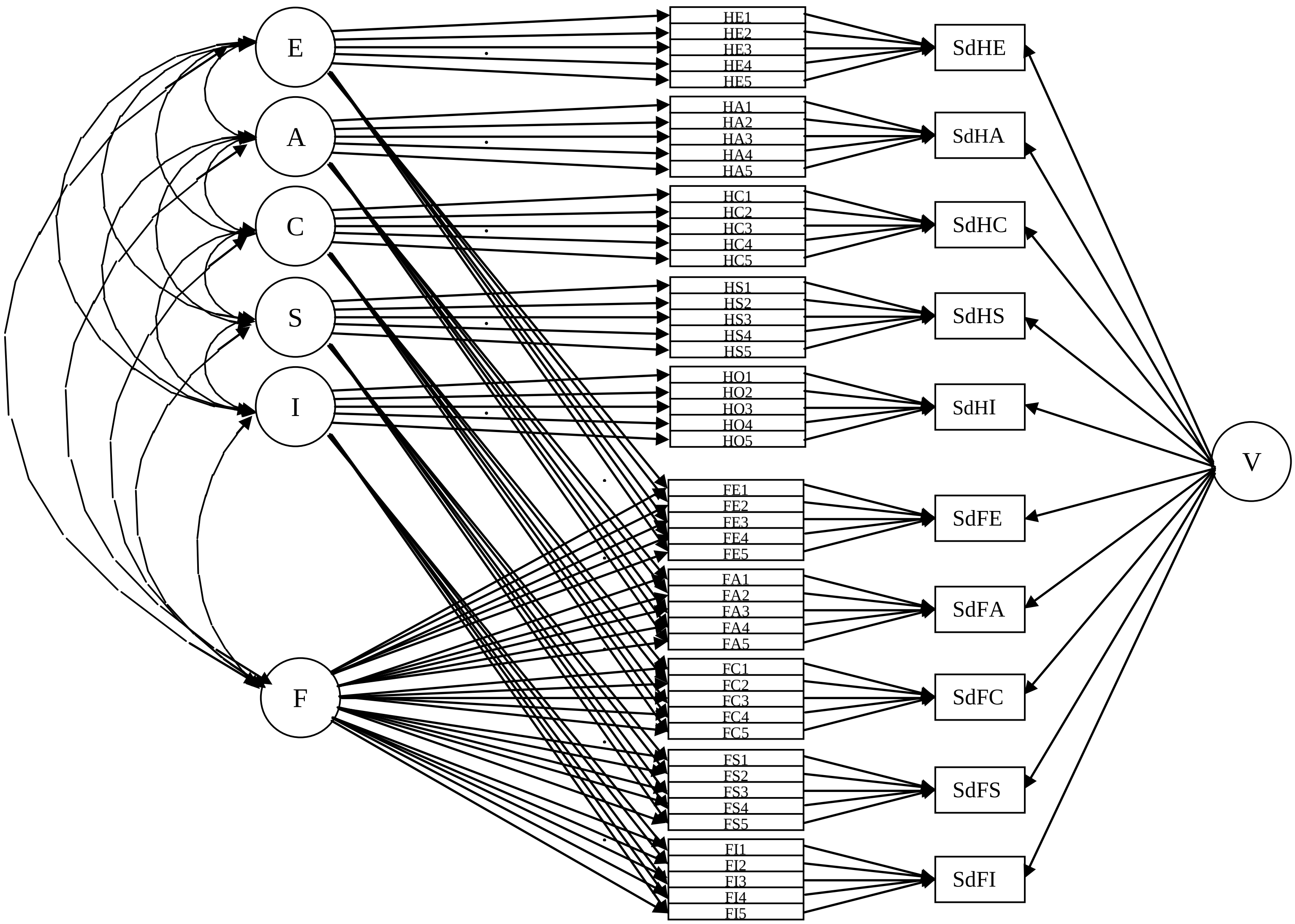


Figure 6. Complete variability model applied to the Clark and Biderman (2006) data. Although not shown in the figure, V was allowed to correlate with FP, FA, and the Big Five latent variables.

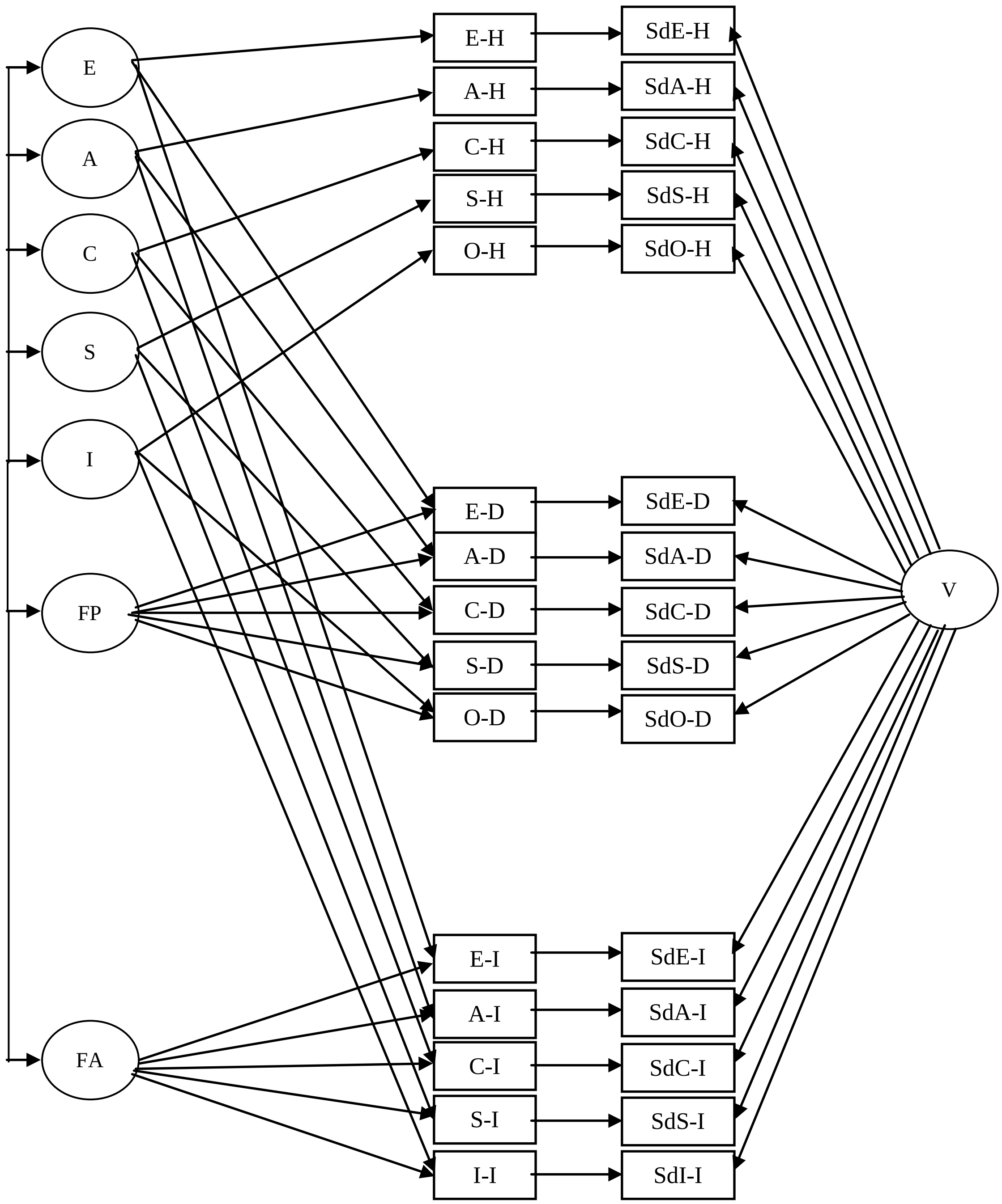
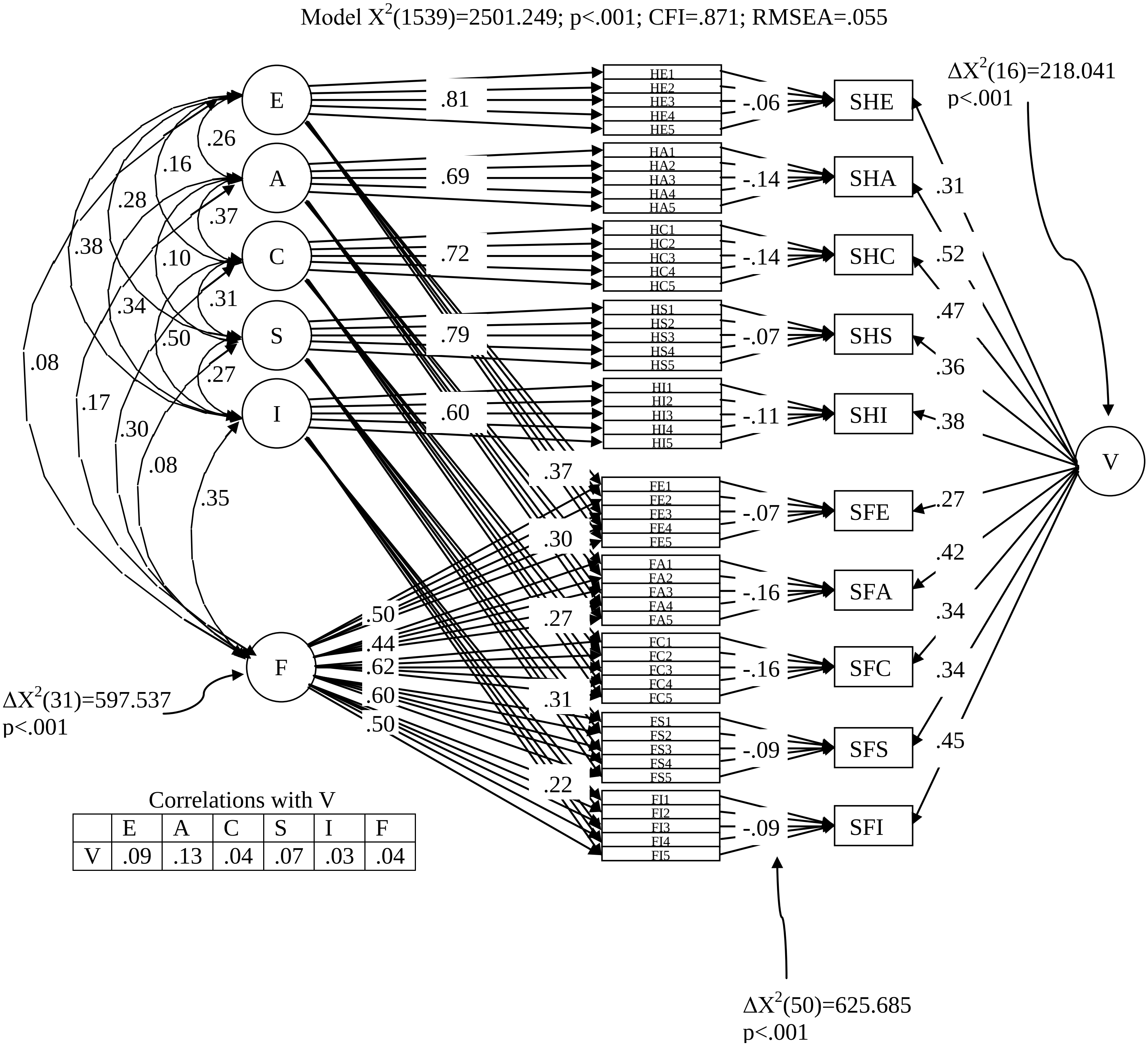


Figure 7. Results of application of the Variability model to the Biderman and Nguyen (2004) five-parcel data. For loadings on Big Five and F latent variables and for regressions of standard deviations onto parcels, means of doubly standardized regression weights for each dimension are presented. Correlations of V latent variable with other latent variables are not represented in the path diagram but are presented in a table at the bottom of the figure.

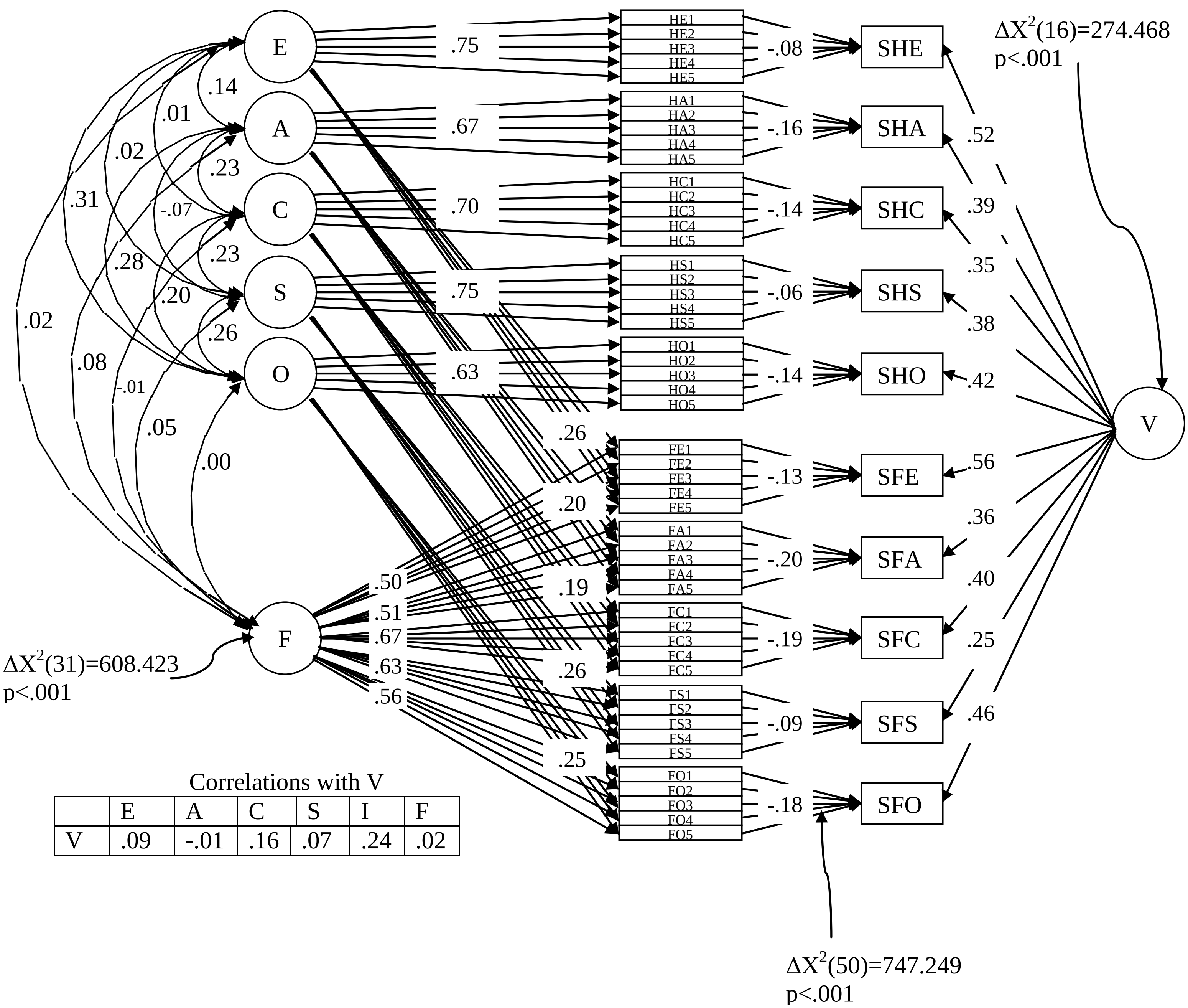


Correlations with V

	E	A	C	S	I	F
V	.09	.13	.04	.07	.03	.04

Figure 8. Results of application of the Variability model to the Wrensen and Biderman (2005) five-parcel data. For loadings on Big Five and F latent variables and for regressions of standard deviations onto parcels, means of doubly standardized regression weights for each dimension are presented. Correlations of V latent variable with other latent variables are not represented in the path diagram but are presented in a table at the bottom of the figure.

Model $X^2(1539)=2666.874$; $p<.001$; CFI=.816; RMSEA=.066

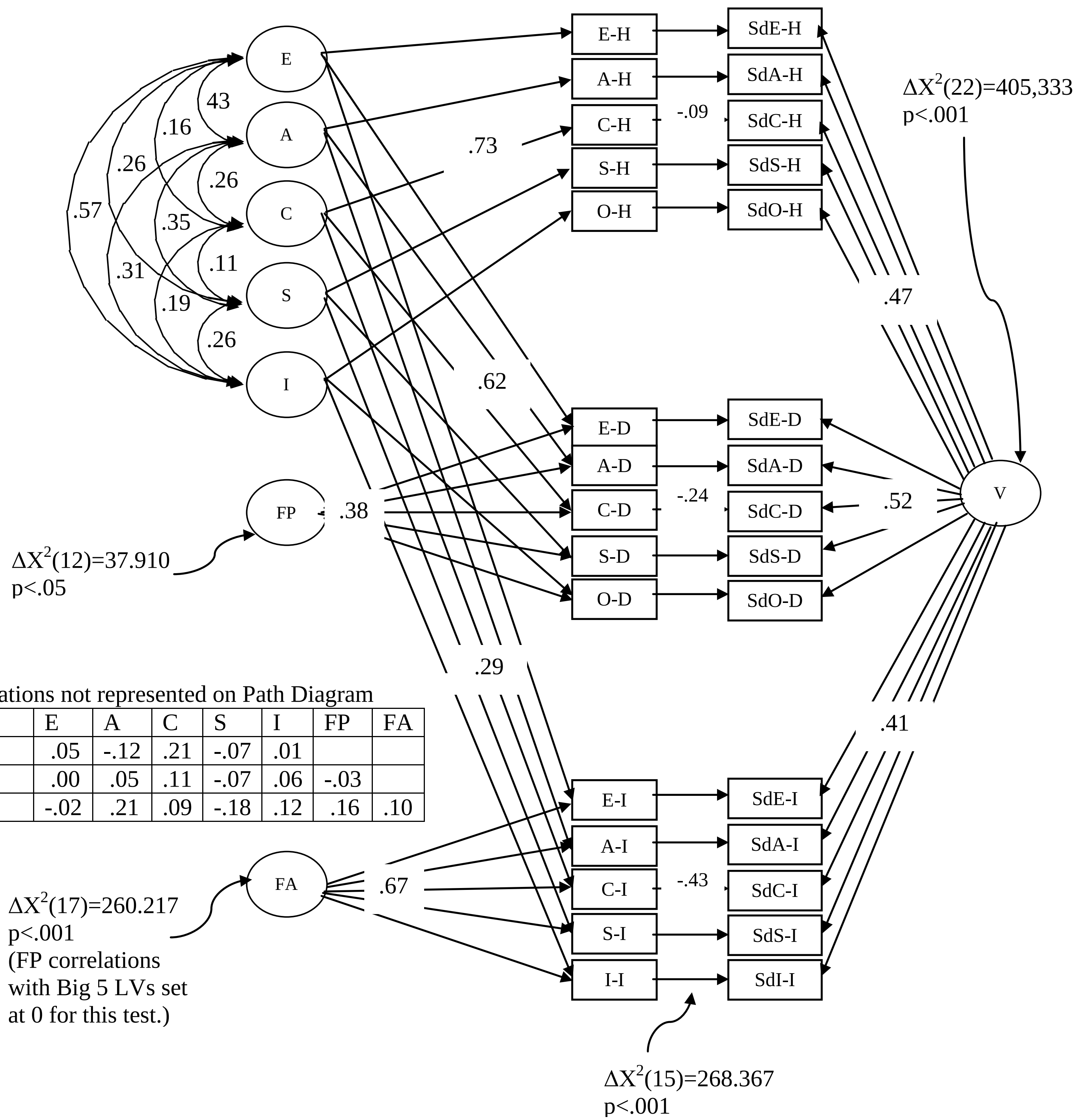


Correlations with V

	E	A	C	S	I	F
V	.09	-.01	.16	.07	.24	.02

Figure 9. Results of application of the Variability model to the Clark and Biderman (2006) whole-scale-score- data. Correlations of the V, FP, and FA latent variable with other latent variables are not represented in the path diagram but are presented in a table at the bottom of the figure.

Model $X^2(352)=532.552$; $p<.001$; CFI=.883; RMSEA=.056



Correlations not represented on Path Diagram

	E	A	C	S	I	FP	FA
FP	.05	-.12	.21	-.07	.01		
FA	.00	.05	.11	-.07	.06	-.03	
V	-.02	.21	.09	-.18	.12	.16	.10