Common method variance in NEO-FFI and IPIP personality measurement

Michael D. Biderman
University of Tennessee at Chattanooga

Nhung T. Nguyen
Towson University

Christopher J. L. Cunningham
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

Poster
TITLE
Common method variance in NEO-FFI and IPIP personality measurement

ABSTRACT
Including a method factor on which all items loaded significantly improved fit of CFAs of both IPIP and NEO-FFI data. Fit was further improved by separating the method factor into two factors representing construct-positive and construct-negative items. The criterion-related validity of conscientiousness increased when uncontaminated by method bias.

PRESS PARAGRAPH
This research extends previous findings of contamination of personality scores by common method bias to the popular NEO-FFI measure of personality traits. Previous studies have found that responses to the IPIP Five Factor Model personality assessment are contaminated by common method bias, and that factors representing construct-positive and construct-negative items contribute separately to that contamination. The present research found similar results for the NEO-FFI measure, indicating almost 14 percent of total variance in responses is associated with such response biases. The effects of this contamination on validity are shown and suggestions for dealing with it are discussed.

Despite widespread use and acceptance of personality assessments such as the NEO-PI (Costa & McCrae, 1985) over the past twenty years, a recent review questions the appropriateness and utility of such assessments for employment selection and other high-stakes testing situations (Morgeson, Campion, Dipboye, Murphy, & Schmitt, 2007). A major criticism of personality assessment raised by these and other researchers is that such assessments rely predominantly on self-reported information. In such assessments, applicants are asked to rate their agreement with multiple statements of behavioral descriptions that supposedly underlie specific personality constructs (i.e., "I have little concern of others"; "I have a good imagination").

A potential consequence of relying on this type of self-reported information is that the resulting scores may be influenced by variance stemming from response tendencies and/or the valences of the items – whether they were positively related to the construct of interest or negatively related to it – rather than the actual target personality construct dimensions alone. More specifically, the issue is that the observed covariance between variables of interest could be inflated or deflated by variance due to the assessment method rather than to the underlying constructs or variables of interest (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

The most commonly used model of personality is the five factor model (FFM) consisting of extraversion, agreeableness, conscientiousness, neuroticism (often measured as emotional stability), and openness to experience (sometimes called intellect or culture). The potential for measures of the FFM traits to be influenced by common method variance was first reported by Schmit and Ryan (1993) who factor analyzed responses to individual items of the NEO-Five Factor Inventory (FFI), a short form of the NEO-Personality Inventory (PI; Costa & McCrae, 1985), within applicant and non-applicant samples in an organization. An exploratory factor analysis (EFA) of a non-applicant sample demonstrated the expected five-factor solution, but in the applicant sample, a six-factor solution fit the data best. Schmit and Ryan labeled this sixth factor an "ideal employee" factor, noting that it, "included a conglomerate of item composites from across four of the five subscales of the NEO-FFI" (Schmit & Ryan, 1993, p. 971). Interestingly, items from all five NEO-FFI subscales loaded on this factor, suggesting that the "ideal employee factor" represented a form of common method bias.

Cellar, Miller, Doverspike, and Klawsky (1996) used confirmatory factor analysis (CFA) to estimate a sixth factor on which composite indicators from all of the FFM dimensions of the long form of the NEO-PI were required to load. The authors found that adding the sixth factor significantly improved model fit of the NEO-PI data. Similarly, Lim and Ployhart (2006) in a multi-trait, multi-method (MTMM) analysis of scale scores found that adding two orthogonal method factors, one associated with NEO-FFI data and one associated with data from a second common personality assessment, the International Personality Item Pool (IPIP; Goldberg, 1999), improved model fit for both data sets.

Item Wording Effects

As discussed, the use of self-report questionnaires to measure personality is a common practice. Conventional wisdom suggests that to reduce response biases such as acquiescence (Nunnally, 1978), it is necessary in scale development to include an equal number of construct-positive or standard scored items, which reflect positively on a participant's standing on a particular construct (e.g., "I am life of the party" is a

construct-positive item for Extraversion) and construct-negative or reverse scored items, which are negatively associated with a particular construct (e.g., "I don't talk a lot"). In assessing Extraversion, for example, if a five-point response scale of agreement is used, then a "5" response on a construct-positive item (e.g., I am the life of the party.) should represent roughly the same amount of Extraversion as "1" for a construct-negative item (e.g., I don't talk a lot). It is standard to assume that large positive response values over an entire set of items represents a greater amount of whatever construct is being measured regardless of the underlying item wording.

The practice of using positive and negative item wording formats to counteract respondents' acquiescence can be found throughout most areas of organizational research including personality assessment (e.g., Paulhus, 1991; Motl & DiStefano, 2002; Quilty et al., 2006), leadership behavior (e.g., Schriesheim & Eisenbach, 1995), role stress (Rizzo, House, & Lirtzman, 1970), and organizational commitment (e.g., Meyer & Allen, 1984).

Unfortunately, the construct-negative items that were introduced to counter response biases such as acquiescence have been found to be associated with systematic and construct-irrelevant variance in scale scores. For example, Hensley and Roberts (1976) conducted an EFA of Rosenberg's Self-esteem scale and found that the scale consisted of two factors: one loading on construct-positive and one on construct-negative items. This finding was later replicated and the factors labeled positive and negative self-esteem (Carmines & Zeller, 1979). Additional studies using CFAs all showed that a model in which the same two method effects were estimated provided the best fit to the data (e.g., Marsh, 1996; Tómas & Oliver, 1999).

Although the existence of common method variance and variance due to item wording has been recognized for many years, there is little evidence regarding the extent to which this type of variance influences scores obtained from common personality assessments. This is an important issue given the widespread use of these types of assessments for research and practice through many subfields of psychology and management. Two of the most popular such assessments are the NEO-FFI (Costa & McCrea, 1989) and the IPIP scales (Goldberg, 1999). While the NEO-FFI is copyrighted and only available at a cost, the IPIP scales are freely available as downloads from the IPIP web site. For this reason, the IPIP scales are very heavily used by researchers needing an assessment of FFM traits.

Recently, Nguyen, Biderman, Cunningham, and Ghorbani (2008) investigated the issue of common method variance by applying a series of CFA models to IPIP 50-item FFM data obtained from seven U.S. and non U.S. samples. In these models a common method factor was estimated on which all items in addition to the five *a priori* factors loaded. In all datasets the model fit was significantly improved when the sixth factor, a single latent variable similar to what Podsakoff and colleagues (2003) labeled an "unmeasured method" effect, was estimated.

Nguyen et al. (2008) also estimated models in which the single method factor was replaced by two factors – one on which all construct-positive items loaded and one on which all construct-negative items loaded. The two-method-factor model significantly improved goodness-of-fit for all seven datasets. Nguyen et al. reported that on average, 14.1% of total variance was explained by these method factors across seven samples. Thus, these results provide compelling and substantial evidence that responses to the IPIP questionnaire are contaminated by common method variance.

To the best of our knowledge, no prior research has looked at the potential influence of item scoring on the factor structure of the NEO-FFI. In addition, there have been few, if any comparisons of the influence of such bias on the predictive validity of these two widely used assessments. Thus, the overarching purpose of the present study was to investigate the presence and implications of such contamination in the more widely used NEO-FFI questionnaire. Further, studies reviewed earlier in the paper that examined the factor structure of the NEO-FFI or NEO-PI only achieved reasonably good fit using parcels as indicators in confirmatory factor analyses. The present paper reports the results of the first examination of this item scoring bias in the NEO-FFI using individual items as indicators.

We had three specific purposes in this study. First, we examined the influence of common method variance on the factor structure of the NEO-FFI. Given the fact that method variance was found to exist in IPIP data across seven samples (Nguyen et al., 2008), we expected that method variance would also exist in NEO-FFI data such that adding a method factor would significantly explain NEO-FFI data. Second, we examined whether item wording accounts for item covariance in NEO-FFI data. As Nguyen et al. (2008) found that factors representing construct-positive and construct-negative items accounted for 14% of the variance across seven datasets, we expected adding two item wording factors would yield similar results in the NEO-FFI data.

Third, we examined the validity of conscientiousness scores for predicting UGPA. The scores from both the IPIP and NEO-FFI were computed without taking into account method variance and computed taking such variance into account. Biderman, Nguyen, and Sebren (2008) found that method bias associated with the FFM questionnaire served as a suppressant of the relationship of conscientiousness to test performance. When purer estimates of conscientiousness were created by statistically removing contamination due to method bias, conscientiousness became a significant predictor of academic performance. Based on the above discussion, we expected that conscientiousness would be a better predictor of academic performance when uncontaminated by method bias and/or item wording biases.

METHOD

Participants

Participants were 197 students at a mid-sized Southeastern university in the U.S. All participated for extra credit in courses.

Procedure

Participants participated in groups limited by the size of the classroom chosen. The questionnaires were distributed and participants filled them out under no external time pressure. Order of the FFM questionnaires within the packet was counterbalanced.

Measures

FFM traits. The Sample 50-item scale available on the IPIP web site (Goldberg, 1999) was used. The NEO responses were entered in a rectangular response-area on the second page of items by blackening the appropriate circle from five response options distributed horizontally with the labels SD, D, N, A, and SA (Strongly Disagree to Strongly Agree). The response block allowed the entry of five responses per line. A similar-appearing response block was created for the IPIP and place at the end of the second page of IPIP items.

The NEO-FFI was developed by Costa and McCrae as a briefer alternative to the 240-item NEO-PI . The NEO-FFI contains 60 items with 12 items measuring each of the FFM dimensions. Scale scores were computed based on the manual accompanying the test booklets purchased from Psychological Assessment Resources, Inc.

Academic performance. Participants gave permission for investigators to access their student records, including grade-point average (UGPA) and ACT composite scores. UGPAs recorded through the semester prior to that in which participation occurred were used. UGPA was not available for those students in their first semester. For participants with only SAT scores, estimated ACT composite scores were created from the SAT scores using an SAT-ACT concordance table (http://professionals.collegeboard.com).

Analysis Process

Three CFAs were conducted for both personality assessments. In the first, only the FFM latent variables were estimated. In the second, a single latent variable was added to the model on which all items loaded. This model is shown in Figure 1. In the third model, the single added latent variable was replaced with two, one on which all construct-positive items loaded and one on which all construct-negative items loaded (see Figure 2). All bias latent variables were estimated as orthogonal to the FFM latent variables. Individual items were used as indicators in all models.

RESULTS

Table 1 presents the descriptive statistics and correlations of scale scores assessed by the NEO-FFI and IPIP. The monotrait-hetero method correlations between the NEO-FFI and IPIP scales were positive and quite high (mean $r = .62$, ranging from .37 to .86). This finding provides evidence of convergent validity of the NEO-FFI and IPIP. The mean heterotrait-monomethod correlation for the IPIP was much lower at .15, while for the NEO-FFI it was .18. The mean heterotrait-heteromethod correlation was .11. Taken together, our results provide evidence for both discriminant and convergent validity of the two personality assessments in line with Campbell and Fiske's (1959) criteria and are consistent with Lim and Ployhart's (2006) study.

Table 2 presents goodness-of-fit statistics for the three CFAs applied to both personality assessments. As can be seen from this table, the model with a single method factor fit significantly better than the model without a method factor for both measures. Moreover, the model with method factors for construct-positive and construct-negative items fit significantly better than the model with just one such factor. Table 3 presents mean loadings from the two questionnaires for the two-method-factor model. In general, all loadings were positive and significant. Since the FFM and method factor latent variables were estimated as orthogonal, the proportion of variance due to item wording factors could be computed. For the IPIP scales, 13.75% of total variance was explained by the item wording factors, comparable to the value of 14.1% reported by Nguyen et al. (2008). For the NEO-FFI, the corresponding value was 13.70%.

For validity analyses, factor scores from the two-method-bias model were computed for each questionnaire and added to the dataset containing summated scale scores and UGPA values. In order to maximize the stability of the UGPAs only participants with more than 15 hours of completed coursework were included in the analyses. The top panel of Table 4 presents the validity results for both conscientiousness represented as summated scale scores and conscientiousness represented as factor scores. As seen from the table, for the IPIP, the correlation between conscientiousness and

UGPA increased from ($r = .18$, $p < .05$) to ($r = .28$, $p < .01$) when conscientiousness was estimated as a summated scale score versus a factor score without contamination of item wording bias (t for correlated rs =3.59, p < .001). Likewise, for the NEO-FFI, the correlation of conscientiousness with UGPA was near zero and non-significant ($r = .03$) when it was represented by a summated scale score but became statistically significant ($r = .22$, $p < .05$) when it was represented by factor scores (t for correlated rs = 2.56, p < .05).

A similar analysis was performed to examine the incremental validity of conscientiousness scales and factor scores controlling for ACT scores. The results of that analysis are presented in the bottom panel of Table 4. The results parallel the simple validity analyses in that while the incremental validities of both IPIP and NEO conscientiousness scale scores were not significant, those of the factor scores, in which method variance had been removed, significantly increased validity for both questionnaires.

DISCUSSION

The purpose of our present study was to investigate method bias in the NEO-FFI. Using sums of squared loadings across items, the magnitude of bias due to item wording factors was estimated to be 13.75% of total variance for the IPIP questionnaire and 13.70% for the NEO-FFI. The results suggest that summated scale scores of the latter questionnaire are also contaminated by common method bias as well as item wording bias. Our results are consistent with previous studies (e.g., Cote & Buckley, 1987; Doty & Glick, 1998) although the total percentage of variance due to method bias was smaller here. We note, though, that the percentage of *explained* variance in the present studies was 33 for the IPIP scales and 40 for the NEO scales, values comparable to the 39 and 41 percent of explained variance due to method reported by Cote and Buckley (1987), and Dote and Glick (1998), respectively. We attribute the difference in percentage of *total* variance accounted for to the fact that our analyses used individual items as indicators while the majority of previous MTMM studies have used scale scores. Because using items increases the possibility of unmodeled covariances between indicators, the greater number of indicators for a given model results in a greater percentage of error variance.

Our research extends the extant literature in at least three ways. First, it extends previous finding of method variance in the NEO-FFI (Schmit & Ryan, 1993; Lim & Ployhart, 2006) by showing that the method bias exists at the item level among respondents with little or no motivation to "fake good". Schmit and Ryan (1993) found evidence for method bias only among applicants, while Lim and Ployhart (2006) identified method variance only in analyses of whole-scale scores. The results here suggest that whatever bias is operating is one that affects virtually all items on a questionnaire regardless of dimension and does not require a faking incentive or instruction.

Second, this study represents the first attempt to document method bias specific to item wording in the NEO-FFI. Specifically we found biases that depend on whether items are positive or negative with respect to the construct of interest. An implication of this finding here and the previous finding by Nguyen et al. (2008), is that CFAs of FFM data that do not estimate item wording factors will likely be misspecified. This is important to our understanding of the processes underlying responses to personality assessments because modeling at the item level allows identification (using item loadings) of those

items most strongly influenced by item wording, content, and/or common method. The method variance found here also implies that FFM scale scores will be contaminated by such variance, resulting in biases in the correlations of those scale scores with other variables of interest.

Third and lastly, the present study is the first to examine the effect of method factors representing item wording on the criterion-related validity of personality measures. This study replicates and extends the finding of Biderman et al. (2008) in which the validity of conscientiousness as a predictor of test performance was increased when it was measured by factor scores from a one method factor model. The present study extended their results to a different academic criterion with conscientiousness estimated from a two method factor model applied to two different personality questionnaires. Specifically, when scale scores were analyzed, only the IPIP conscientiousness scale correlated significantly with UGPA in a simple validity analysis. However, when factor scores were estimated from the two method factor model, those factor scores from both the IPIP and the NEO-FFI were significantly related to UGPA in *both* simple validity and incremental validity analyses, as shown in Table 4. These results clearly indicate that the presence of method variance, even in one of the variables in a predictor-criterion relationship may very likely suppress that relationship.

The results of the present study along with those of the Nguyen et al. (2008) are remarkably consistent regarding the influence of method variance on responses to personality assessments. The existence of these findings calls for speculation and further research on the source of the variance and for factors that affect its magnitude. The study of Schmit and Ryan (1993) and subsequent reports by others, e.g., Biderman and Nguyen (2004) suggest that incentives or instructions to fake may increase method variance, although whether method variance under instruction to respond honestly has the same source as method variance under conditions with incentives or instructions to fake remains to be seen. Moreover, it is doubtful that the item wording factors found consistently in these datasets could be the result of faking alone.

Although there were significant differences in the fit of the two-method-bias and that of the single bias factor model in this study and in the seven datasets reported by Nguyen et al. (2008), the correlations between the two method bias factors were quite high in several of the datasets with five of the seven being larger than .75 and two being less than .50 in the (Nguyen et al., 2008) study. In the present study, the correlations between the positive and negative bias factors were .78 for the IPIP and .83 for the NEO-FFI. This suggests that while for most studies, the two wording factors are quite similar, there may be factors that moderate the correlation between the two wording-specific factors.

The implications of the results of this study for personality researchers are clear. These results imply that scale scores for FFM personality assessments are contaminated and that CFAs that do not take method biases into account are misspecified. This appears to be the case even for research involving a single personality dimension. A solution to this problem may involve administration of all FFM dimensions and estimation of method factors, even though scores on only one dimension are used. Researchers desiring estimates for individual respondents may resort to the use of factor scores, such as those employed in the validity analyses of the present study.

CONCLUSIONS

In this study, we showed that method bias and item scoring biases were responsible for variation in item responses in IPIP and NEO-FFI measures under the situation in which participants had little or no motivation to look good. Overall, a model with two method bias factors – one indicated by regular scored and one indicated by reverse scored items fit data the best. Finally, conscientiousness factor scores estimated from the model with two method bias factors were found to have greater validity than scale scores in both simple validity analyses and incremental validity analyses controlling for ACT scores.

References

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

Biderman, M. D., Nguyen, N. T., & Sebren, J. (2008). Time on task mediates the conscientiousness-performance relationship. *Personality and Individual Differences,44*, 887-897.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment.* Newbury Park, CA: Sage.

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

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

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

Cote, J., & Buckley, M. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24,* 315-318.

Doty, D. H., & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods, 1,* 374–406.

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.

Hensley, W. E., & Roberts, M. K. (1976). Dimensions of Rosenberg's Self-esteem scale. *Psychological Reports, 78*: 1071-1074.

Lim, B-C., & Ployhart, R. E. (2006). Assessing the Convergent and Discriminant Validity of Goldberg's International Personality Item Pool: A Multitrait-Multimethod Examination. *Organizational Research Methods, 9*, 29-54.

Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or arfactors? *Journal of Personality and Social Psychology, 70*: 810-819.

Meyer, J., & Allen, N. (1984). Testing the "Side-bet theory" of organizational commitment: Some methodological considerations. *Journal of Applied Psychology, 69*: 372-378.

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

Motl, R. W., & DeStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling, 9*, 562-578.

Nguyen, N.T., Biderman, M.D., Cunningham, C. L., & Ghorbani, N, (2008). *A Multi-Sample Re-examination of the Factor Structure of Goldberg's IPIP 50-item Big Five*

Questionnaire. Paper presented at the 2008 Annual Meeting of the Academy of Management: Anaheim, CA.

Nunnally, J.C. (1978). *Psychometric theory*, 2nd ed. New York: McGraw-Hill.

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

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.

Rizzo, J. R., House, R. J., & Lirtzman, S. I. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly, 15*: 150-163.

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.

Schriesheim, C.A., & Eisenbach, R.J. (1995). An exploratory and confirmatory factor analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *Journal of Management, 21*, 1177-1193.

Tomás, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling, 6*, 84-98.

Table 1. Means, standard deviations, and correlations between measures. Reliability estimates are in parentheses.

Dimension	1	2	3	4	5	6	7	8	9	10
1. IPIP E	(.89)									
2. IPIP A	.31	(.81)								
3. IPIP C	-.01	.28	(.84)							
4. IPIP S	.29	.20	.19	(.89)						
5. IPIP O	.06	.07	.06	-.01	(.78)					
6. NEO E	.70	.48	.24	.43	.09	(.78)				
7. NEO A	.13	.59	.30	.39	-.08	.39	(.79)			
8. NEO C	.00	.07	.37	.06	-.07	.14	.03	(.85)		
9. NEO S	.30	.08	.19	.86	.01	.39	.30	.08	(.89)	
10. NEO O	.00	.04	-.28	.02	.58	-.08	-.08	-.29	.-03	(.78)
Mean	3.37	4.06	3.60	3.16	3.75	3.55	3.66	3.48	3.27	3.24
SD	0.74	0.50	0.61	0.75	0.50	0.52	0.55	0.27	0.74	0.53

Table 2. Goodness of fit of the models.

	Model		
Statistic	No Method Factor	One Method Factor	Two method Factors
IPIP 50-item Scale			
χ^2	2561.018 [a]	2232.979[a]	2201.329[a]
CFI	.669	.735	.743
RMSEA	.078	.072	.071
$\Delta\chi^2$ re previous model		328.039[a]	31.65[a]
Model df	1165	1115	1114
Δdf re previous model		50	1
NEO-FFI			
χ^2	3223.910[a]	2897.786[a]	2843.860[a]
CFI	.639	.702	.715
RMSEA	.068	.063	.061
$\Delta\chi^2$ re previous model		326.124 [a]	53.926[a]
Model df	1700	1640	1639
Δdf re previous model		60	1

[a] $p < .001$

Table 3. Mean loadings from the two-method-factor model for each questionnaire.

Standard deviations are in parentheses.

Latent Variable	IPIP 50-item Scale	NEO-FFI
Extraversion (E)	0.62 (0.09)	0.22 (0.27)
Agreeableness (A)	0.51 (0.16)	0.30 (0.18)
Conscientiousness (C)	0.56 (0.0)	0.53 (0.12)
Stability (S)	0.29 (0.26)	0.48 (0.10)
Openness (O)	0.51 (0.13)	0.47 (0.22)
Mp – Positively-worded item factor	0.22 (0.21)	0.30 (0.19)
Mn – Negatively-worded item factor	0.36 (0.23)	0. 31 (0.23)

Table 4. Validity analyses of conscientiousness predicting UGPA.

Simple validity analyses (N=117)

	IPIP Questionnaire		NEO Questionnaire	
	Scale Scores	Factor Scores	Scale Scores	Factor Scores
UGPA	.183[a]	.282[b]	.029	.224[a]

Incremental validity analyses (N=97)

	IPIP scale Scores	IPIP Factor Scores
Predictor	Standardized Coefficient	Standardized Coefficient
ACT comp	.544	.562[c]
C	.127	.196[a]

	NEO scale Scores	NEO Factor Scores
Predictor	Standardized Coefficient	Standardized Coefficient
ACT comp	.556	.535[c]
C	.087	.183[a]

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

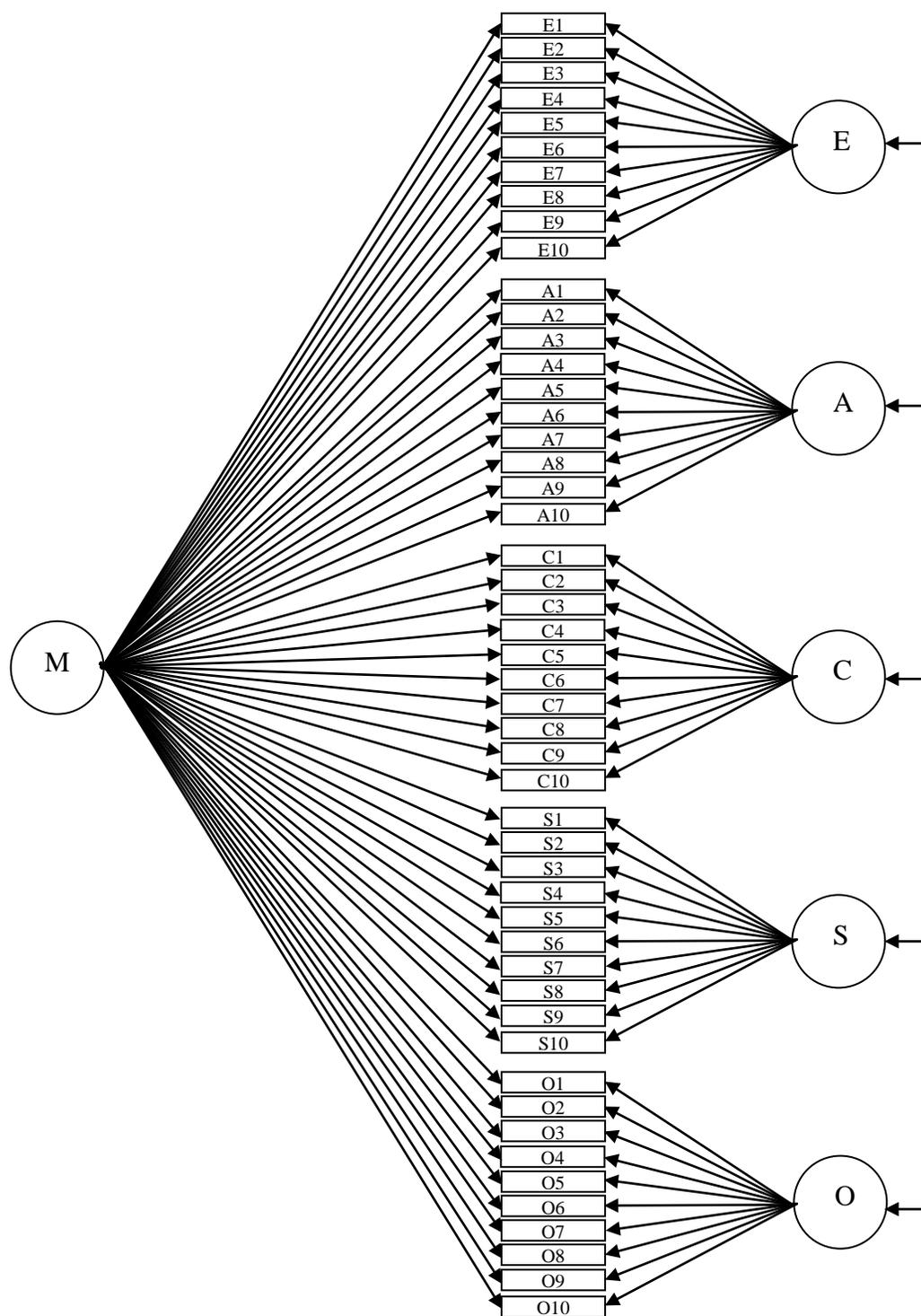Figure 1. CFA with one method bias factor as applied to the IPIP 50-item questionnaire.

Figure 2. CFA with two method-bias factors, one indicated by positively-worded items and one indicted by negatively-worded items as applied to the IPIP questionnaire.