

Method factors, bifactors, and item valence

Michael D. Biderman

University of Tennessee at Chattanooga

Nhung T. Nguyen

Towson University

Christopher J. L. Cunningham

Zhuo Chen

Paul J. Watson

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](mailto:Michael-Biderman@utc.edu)

Paper accepted for presentation at the 28th Annual Conference of The Society for Industrial and Organizational Psychology, Houston, TX 2013.

Poster

TITLE

Method factors, bifactors, and item valence

ABSTRACT

A method factor/bifactor model fit Big Five data significantly better than a simple CFA model. Item loadings on the bifactor were positively related to independently gathered ratings of item valence. Results are consistent with the hypothesis that there is a common affective component in responses to all Big Five items.

PRESS PARAGRAPH

Application of a method or bifactor confirmatory factor analysis model of responses to a Big Five personality questionnaire revealed that the valence of individual items influenced response biases to the items. Positive response biases were greatest for those items judged to describe very good characteristics and negative biases were greatest for those judged to describe very bad characteristics. The results here combined with previous results suggest that responses to all personality items are based partially on the affective state of the respondent.

The factor structure of personality measured by Big Five questionnaires has become a cornerstone of research since first being popularized in the 1990s (Costa & McCrae, 1992; Digman, 1990; Goldberg, 1993). The original intent of investigators in this area may have been to operationalize orthogonal personality dimensions, but in practice, scale or domain scores of the Big Five dimensions – Neuroticism/Stability, Extraversion, Openness, Conscientiousness, and Agreeableness – have often correlated. When Neuroticism is reverse-scored and labeled Stability, as in the present project, these correlations are all generally positive; forming what is termed a positive manifold (e.g., Paulhus & Martin, 1987).

One possible explanation for the positive manifold is the existence of a common influence on all the items. Such common influences have frequently been studied as method factors – influences on behavior associated with the use of a common method for collection of all responses, e.g., self-report. The assumption is that there are individual differences in response biases that apply to all items in a self-report questionnaire, and these individual differences result in positive correlations among all the items. When the items are aggregated into scale scores, the positive correlations among items in different scales result in positive correlations among the scales – the positive manifold. Biderman, Nguyen, Cunningham, and Ghorbani (2011) provided evidence for the existence of such common method factors for two Big Five questionnaires – the IPIP 50-item sample questionnaire (Goldberg, 1999; [www.ipip.org](http://www.ipip.org)) and for the NEO-FFI questionnaire (Costa & McCrae, 1989). In fact, Biderman et al. (2011) found evidence for three common factors affecting items responses – a general factor affecting responses to all items, a factor affecting only positively worded items, and a factor affecting only negatively worded items. The focus of the present study is on the general factor.

As noted by Biderman et al. (2011) confirmatory factor analysis (CFA) models of questionnaire responses that incorporate both general factor influencing all items and group factors, each affecting only a subset of items, are called bifactor models (Holzinger & Swineford, 1937). Models incorporating such factors are also called nested factor models (Gustafsson & Balke, 1993) although the term bifactor will be used here. Bifactor models have usefully represented data from several non-Big Five personality assessments (e.g., Brouwer, Meijer, & Weekers, & Baneke, 2008; Chen, Hayes, Carver, Laurenceau, & Zhang, In press; Gignac, Palmer, & Stough, 2007; Reise, Moore, & Haviland, 2010; Reise, Morizot, & Hays, 2007). In these studies, bifactor models represented both the central construct and the group factors more effectively than models in which the central construct was represented by a higher order factor. Recently bifactor models have also been applied to the study of the structure of intelligence, in contrast to the traditionally held higher order factor structure (Drasgow, Nye, Caretta, & Ree, 2010; Gignac, 2007; Gignac, 2005; Gignac, 2006; Golay & Lecerf, 2011; Lang, Kersting, Hulsheger, & Lang, 2010). Once again, the bifactor models fit the data better than did traditional higher order factor models.

In addition to the Biderman et al. (2011) study mentioned above, bifactor models have also been applied to the data of Big Five questionnaires (Bäckström, 2007; Bäckström, Björklund, & Larsson, 2009; Cellar, Doverspike, Miller, & Klosky, 1996; Klehe, Kleinmann, Hartstein, Melchers, König, Heslin, & Lievens, in press). An illustration of a model with a single general bifactor applied to items of a Big Five questionnaire with 10 items per dimension is shown in Figure 1. Two characteristics of such models should be pointed out in the figure. First, the common method / bifactor is a first order factor, not a higher order factor. Its influence is on the primary indicators – the item responses – and not mediated by the other factors. Second, the

primary indicators are items, rather than scale scores. The basic trait/method unit in such a model is the item response. It is not possible to estimate a bifactor model when scale scores are the primary indicators without making unwarranted restrictions on the covariances among the scales.

The present study stemmed from suggestions of Biderman et al.(2011) that the bifactor might reflect responses to the valence or evaluative characteristics of the items (e.g., Benet-Martinez & Waller, 2002; Waller, 1999). Their suggestion was based on the discovery that the bifactor was correlated with measures of affect – positively with positive affect and negatively with negative affect, as measured using the PANAS. This finding was replicated by Biderman, Nguyen, & Cunningham (2012) who found it positively related to self-esteem and negatively related to depression. Based on these results, we wished to determine whether evaluative aspects of individual items were related to the influence of the bifactor. To that end we had respondents evaluate the valence of items in the IPIP Big Five Questionnaire– to indicate how much the characteristic represented by the item would make them look bad or look good. We then correlated those valence estimates with loadings of those items on the bifactor when applied to a data from different sets of respondents. If the Biderman et al. (2011) suggestion was correct, we expected positive correlations, suggesting that the relationships of item responses to the bifactor were determined to some extent by valences of the items, with the strongest relationships among items with the greatest valence.

## **Method**

### **Participants**

To maximize sample size for application of the bifactor model, data from three separate samples were concatenated into a single sample. Although some of the data have been reported in previously published articles, none of these published articles have involved analyses

involving item valence. Sample sizes for the three data sets were 183, 158, and 206, totaling 547. Participants were undergraduate students at mid-sized universities in the United States. Data for application of the bifactor model were gathered between 2004 and 2010. Mean age was 19.97 with standard deviation equal to 4.08. Percentage of males was 36.20. Percentage of respondents designating themselves as White/Caucasian, African-American, and Other were 67.30, 25.10, and 7.60 respectively.

Data for obtaining valence estimates were obtained in the Fall semester of 2011 ( $N = 366$ ) and the Spring semester of 2012 ( $N = 303$ ) from a mid-sized university in the United States. Percentage of males in the combined samples was 39.50. Percentages designating themselves as White/Caucasian, African-American and Other were 12.20, 80.20, and 7.60 respectively.

### Measures

**IPIP Big Five.** The Big Five questionnaires was the 50-item sample questionnaire posted on the web site of the International Personality Item Pool ([www.ipip.org](http://www.ipip.org)). For the data to which models were to be applied, respondents were asked to indicate “how accurately each statement describes you”. Responses occurred on a seven-point scale with one end of the scale labeled “Very accurate” and the other end labeled “Very inaccurate”. All participants responded under instructions to respond honestly with no instructions or incentives to fake.

**Valence Judgments.** In the two samples from which item valence ratings were obtained, respondents received the IPIP 50-item questionnaire and indicated how they would be evaluated by other people displaying the characteristic described by the item. The Fall 2011 sample ( $N=366$ ) responded on a five-point scale, while the Spring 2012 sample ( $N=303$ ) responded on a seven-point scale, with response options ranging from “They would say that if I had this

characteristic, it would make me look very bad” to “They would say that if I had this characteristic, it would make me look very good.”

### **Models**

Models were estimated using the method of maximum likelihood as implemented by *Mplus* V 6.12 (Muthén & Muthén, 1998-2012). Responses to individual items were analyzed. Model 1 was a simple CFA of the 50 items, with one factor for each Big Five dimension. It allowed the factors to covary. Model 2 was the model shown in Figure 1 the same CFA but with an added bifactor on which all items loaded. Measures of goodness-of-fit reported here are the chi-square and chi-square difference statistics, the CFI, TFL, the RMSEA, and the SRMR statistics.

### **Results**

Table 1 presents correlations between Big Five scale scores. The data display a positive manifold - all of the correlations were positive ( $M = .202$ ). Table 2 presents goodness-of-fit statistics for comparisons of the two models. In this comparison all measures of goodness-of-fit suggest that the bifactor model provided a better accounting of the data than did the simple CFA. The large improvement warrants further investigation of the bifactor. Table 3 presents correlations between the Big Five factors from Model 2. The mean of the 10 correlations is  $-.034$ , considerably reduced and closer to zero than the mean of the scale correlations.

The relationships between bifactor loadings and judgments of item valence were assessed as follows. For the two samples that provided valence judgments, the mean valence rating for each item was calculated. As a manipulation check, the mean of the valence ratings of the negatively worded items was compared with the mean for the positively worded items. If respondents were interpreting the instructions properly, the mean valence of the negatively

worded items, all of which indicated behaviors that would make the respondent “look bad” in the wording of the instructions, should be smaller than the mean of the positively worded items. For the fall 2011 sample, the mean of the negatively worded items on a 0-4 scale was 1.31 ( $SD = 0.45$ ) while the mean of the positively worded items was 2.79 ( $SD = 0.30$ ;  $t(48) = 13.68$ ,  $p < .001$ ,  $d = 3.95$ ). For the spring 2012 sample, the mean for negatively worded items on 1-7 scale was 3.61 ( $SD = 0.49$ ) while that for positively worded items was 5.14 ( $SD = 0.37$ ;  $t(48) = 12.49$ ,  $p < .001$ ,  $d = 3.56$ ). These results suggest that participants’ responses to the items were based on the instructions given.

The ratings of the negatively worded items were then reverse scored so that they would correspond to the similarly reverse scored negatively worded items on the questionnaire given the model-application sample. Standardized loadings of the items on the bifactor from the model application sample were correlated with mean valence ratings. For the fall 2011 sample, the correlation was .427 ( $p < .01$ ) and for the spring 2012 sample, it was .325 ( $p < .05$ ), both results suggesting that the influence of the bifactor on the response to an item depends to some extent on the valence of that item, with larger loadings associated with larger mean valence estimates.

To explore the effect of item valence characteristics on the positive manifold of scales, we conducted three separate analyses on the items. First, for each IPIP dimension, the responses to the five items with the smallest loading on the bifactor were averaged to create a scale based on items with small bifactor loadings. Responses to the remaining items were also averaged to create a scale based on items with large bifactor loadings. After creating the scales – five scales based on items with low loadings and five based on items with high loadings - the 10 correlations between each set of five scales were computed. The mean interscale correlation of the small loading scales was .069 while the mean interscale correlation was .170 for the large

loading scales. We then compared these means by treating each correlation as an individual observation, with the low loading scale correlation for two dimensions paired with the high loading scale correlation for the same dimensions. The difference in mean correlations was significant,  $t(9)=3.298, p < .01, d = 1.17$ .

Next, for each dimension, the five items with the smallest mean valence judgment from the fall 2011 sample were averaged to create a scale based on items with small valence judgments. The remaining items were averaged to create a scale based on items with large valence judgments. The mean of the interscale correlations for the low valence scales was .102; that for the high valence scales was .156. This difference was also significant,  $t(9)=2.834, p < .05, d = 0.69$ .

Finally, a similar splitting of the items was done based on the spring 2012 valence judgments. The mean of the small valence items was .133 while that of the large valence items was .162. This difference, although in the expected direction, was not significant,  $t(9)=1.199, p > .05, d=0.41$ . Taken as a whole, these results provide support for the conclusion that item loadings on the bifactor are related to item valence and that through those loadings, valence influences the size of the positive manifold.

### **Discussion**

The results of comparison of goodness-of-fit of the bifactor model with that of a simple CFA replicates past research comparing such models – the bifactor model fit significantly better than a simple CFA. The improvement in goodness-of-fit of the bifactor model suggests that identification of variables related to the bifactor is warranted. Previous studies have found correlations of measures of respondent characteristics - positive and negative affect - with the bifactor (Biderman et al. (2011)). In this study we found that the bifactor is also related to

characteristics of the questionnaire items. Specifically, the valence analyses suggested that the influence of the bifactor on an item – the item’s loading – was related to the valence of the items. Individual differences in participants’ tendencies to positively or negatively bias their response were most salient for items with the most extreme valence values.

These data suggest that ultimately the effect of item valence on bifactor loadings is on interscale correlations. The differences in the positive manifolds of scales composed of items with either small loadings and less extreme valence or large loadings and more extreme valence showed the effects of item valence and bifactor loadings on those correlations. This result extends the observations of Bäckström et al. (2009) who were able to reword the items of a Big Five scale to reduce their loadings on the bifactor in a model identical to that of Model 2 and presumably also reduce interscale correlations. Our data suggest that the valence estimates obtained here could be the basis for such manipulations.

### **Interpretation of the Bifactor**

The bifactor appears to represent individual differences in a tendency to bias responses to all items positively or negatively. The effects of individual differences in this tendency are exacerbated when the items have extreme valence. The correlations between the bifactor and measures of positive and negative affect reported previously suggest that those respondents most likely to bias their responses positively to items with positive valence are also those who rate themselves as having higher self-esteem and lower depression. We believe this means that under conditions in which respondents have been given no specific instructions except the instruction to respond honestly, the bifactor reflects an individual’s *general tendency to report his/her affective state* while responding to personality assessment items.

The importance of item evaluative characteristics has been addressed for many years. Saucier (2002) and Block (1995) commented on the existence of a general evaluative factor in Big Five data. Almagor, Tellegen, and Waller (1995) argued that excluding evaluative terms from the item pools upon which Big Five factor analyses were based may have precluded the emergence of evaluative dimensions, such as the positive valence and negative valence factors found in the Tellegen and Waller's (1987) Big Seven. The results of the present study provide additional support for the hypothesis that evaluative item characteristics affect responses and also provide a way to capture those along with other Big Five characteristics. The bifactor approach used in this study reveals that characteristics reflecting responses to both traditional content and to item evaluative characteristics are estimable concurrently in one model .

### **Implications for Existing Research**

Two studies have found evidence for the criterion-related validity of the bifactor model in Model 3. In the first, Biderman, Nguyen, Mullins, and Luna (2008) found that compared to Big Five dimensions, the bifactor was the only significant predictor of supervisor ratings of performance of personnel in a personal finance company. In the second, Klehe et al. (in press) found that the bifactor significantly predicted ratings of performance in assessment centers. In both of these studies, it may be argued that persons high on the bifactor – those self-reporting, and presumably presenting high levels of positive affect – would be most likely to receive positive ratings of supervisors or raters, while those self-reporting less positive affect would be likely to receive lower ratings.

Finally, the fact that apparently that all item responses are influenced to some extent by the bifactor factor has implications for the measurement of psychological traits. It implies that items which heretofore had been thought to reflect only the respondent's conscientiousness, for

example, must now be viewed as reflecting both the respondent's conscientiousness and his or her affective state at the time. This suggests that all Big Five scale scores must be viewed as being potentially contaminated by affective state or whatever it is that the bifactor represents. This study and the work of Bäckström et al. (2009) has shown that such contamination can be manipulated to a certain extent. But it may be that psychologists will be better served by moving to techniques that include measuring such "contaminants" as the bifactor and studying their relationships to other variables. Such an approach might give us a better understanding of personality than that obtainable by ignoring such factors.

**References.**

- Almagor, M., Tellegen, A., & Waller, N. G. (1995). The big seven model: A cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality and Social Psychology*, *69*, 300-307.
- Bäckström, M. (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment*, *23*, 63-70.
- Bäckström, M., Björklund, F. & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, *43*, 335- 344.
- Benet-Martinez, V. & Waller, N. G. (2002). From adorable to worthless: Implicit and self-report structure of highly evaluative personality descriptors. *European Journal of Personality*, *16*, 1-41.
- Biderman, M. D., Nguyen, N. T., Cunningham, C. J. L., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality*, *45*, 417-429.
- Biderman, M. D., Nguyen, N. T., & Cunningham, C. J. L. (2011). A method factor measure of self-concept. Paper presented at the 23<sup>rd</sup> Annual conference of The Society for Industrial and Organizational Psychology, Chicago, IL.
- Biderman, M. D., Nguyen, N. T., Mullins, B., & Luna, J. (2008). *A method factor predictor of performance ratings*. Paper presented at the 23<sup>rd</sup> Annual conference of The Society for Industrial and Organizational Psychology, San Francisco, CA.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187-215.

- Brouwer, D., Meijer, R. R., Weekers, A. M., & Baneke, J. J. (2008). On the dimensionality of the Dispositional Hope Scale. *Psychological Assessment, 20*, 310-315.
- Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawnsky, 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.
- Chen, F.F., Hayes, A., Carver, C. S., Laurenceau, J., & Zhang, Z. (In press). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. Accepted Article"; doi: 10.1111/j.1467-6494.2011.00739.
- Costa, P.T., Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P.T., & McCrae, R.R. 1989. *The NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.
- Dragow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualifying Test Form S: Analysis and comparison with previous forms. *Military Psychology, 22*, 68-85.
- Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment, 12*, 320-329.
- Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences, 27*, 73-86.

- Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42*, 37-48.
- Gignac, G. E., Palmer, B. R., & Stough, C. (2007). A confirmatory factor analytic investigation of the TAS-20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment, 89*, 247-257.
- Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment, 23*, 143-152.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-34.
- 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.
- Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*, 407-434.
- Holzinger, K. J., & Swineford, R. (1937). The bifactor method. *Psychometrika, 2*, 41-54.
- Klehe, U.-C., Kleinmann, M., Hartstein, T., Melchers, K., König, C., Heslin, P. & Lievens, F. (in press). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance*.
- Lang, J. W. B., Kersting, M., Hülshager, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology, 63*, 595-640.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (6th ed.). Los Angeles, CA:

Muthén & Muthén.

Paulhus, D. L., & Martin, C. L. (1987). The structure of personality capabilities. *Journal of Personality and Social Psychology*, *52*, 354-365.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Befactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544-559.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*, 19-31.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality*, *36*, 1-31.

Tellegen, A., & Waller, N. G. (1987, August). *Re-examining basic dimensions of natural language trait descriptors*. Paper presented at the 95<sup>th</sup> Annual Convention of the American Psychological association, New York, NY.

Waller, N. G. (1999). Evaluating the structure of personality. In C. R. Cloninger (Ed.), *Personality and Psychopathology* (pp. 155-197). Washington, DC: American Psychiatric Press.

Table 1.

*Scale means, standard deviations, and correlations. Reliability coefficients are on the diagonal.*

---

Variable	Mean	SD	1	2	3	4	5	6	7
<hr/>									
IPIP (N=547)									
1 Extraversion	4.58	1.02	.875						
2 Agreeableness	5.25	0.81	.301 <sup>c</sup>	.802					
3 Conscientiousness	4.71	0.85	.089 <sup>a</sup>	.212 <sup>c</sup>	.806				
4 Stability	4.30	0.99	.222 <sup>c</sup>	.118 <sup>b</sup>	.130 <sup>b</sup>	.844			
5 Openness	4.81	0.81	.241 <sup>c</sup>	.291 <sup>c</sup>	.222 <sup>c</sup>	.188 <sup>c</sup>	.802		

---

<sup>a</sup>  $p < .05$     <sup>b</sup>  $p < .01$     <sup>c</sup>  $p < .001$

Table 2.

*Results from application of the confirmatory factor analysis models to items.*

---

Model	Chi-square	df	CFI	TLI	RMSEA	SRMR
1: Simple CFA	3958.723	1165	.716	.701	.066	.081
2: Bifactor	3454.969 <sup>c</sup>	1115	.763	.740	.062	.066
Difference	503.754	50	.047	.039	.004	.015

---

<sup>a</sup>  $p < .05$     <sup>b</sup>  $p < .01$     <sup>c</sup>  $p < .001$

Table 3.

*Correlations between factors from Model 3. Diagonal entries are factor determinacies.*

---

	E	A	C	S	O
	---	---	---	---	---
<b>IPIP</b>					
E	.869				
A	-.141 <sup>a</sup>	.862			
C	-.181 <sup>b</sup>	-.004	.886		
S	.011	-.282 <sup>c</sup>	-.051	.916	
O	.049	.045	.151 <sup>b</sup>	.066	.881

---

<sup>a</sup>  $p < .05$       <sup>b</sup>  $p < .01$       <sup>c</sup>  $p < .001$

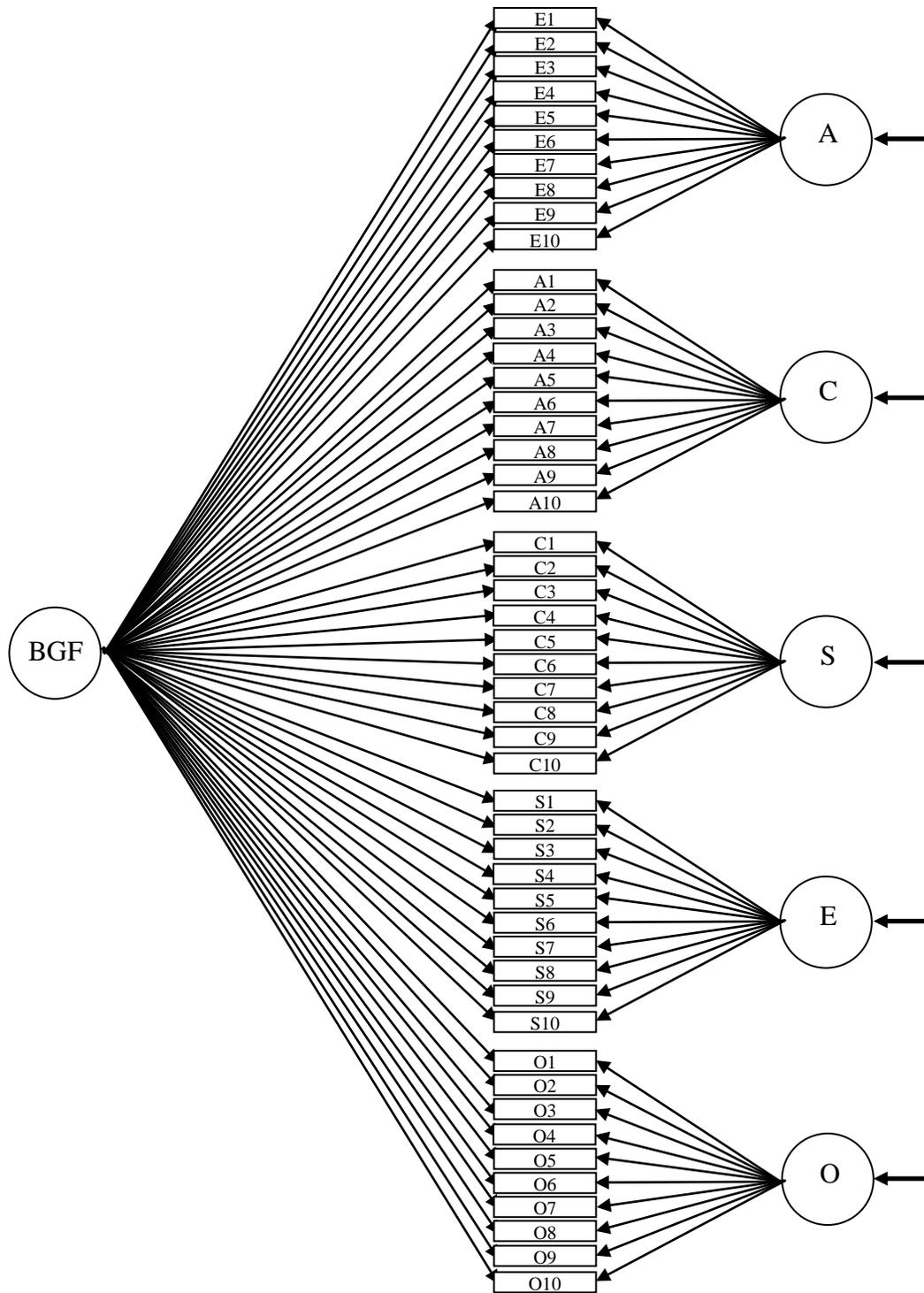


Figure 1. Bifactor Model applied to items of a 50-item Big Five questionnaire. Item residuals omitted for clarity. BGF stands for bifactor general factor.