



## Brief Report

## Random intercept EFA of personality scales



Julian Aichholzer\*

Department of Methods in the Social Sciences, University of Vienna, Rathausstraße 19, 1010 Vienna, Austria

## ARTICLE INFO

## Article history:

Available online 11 July 2014

## Keywords:

Factor analysis  
ESEM  
Simple structure  
Acquiescence  
Big Five  
NEO-FFI  
BFI

## ABSTRACT

Previous research suggests that simple structure CFAs of Big Five personality measures fail to accurately reflect the scale's complex factorial structure, whereas EFAs generally perform better. Another strand of research suggests that acquiescence or uniform response bias masks the scale's "true" factorial structure. Random Intercept EFA (RI-EFA) captures acquiescence as well as the complex item-factor structure typical for personality measures. It is applied to the NEO-FFI and the BFI scale to test whether an accurate model-to-data fit can be achieved and whether the "clarity" of the factorial structure improves. The results lend confidence in the general effectiveness of RI-EFA whenever acquiescence bias is an issue. Example *Mplus* code is provided for replication.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## 1. Introduction

The evaluation of the structure and validity of new and existing measures of psychological traits or attitudes is a central task in quantitative empirical research. One of the most important statistical techniques in this area is the common factor model and its variants, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). In particular, this study focuses on assessing the factorial structure of "Big Five" personality trait measures by means of factor analysis.

First, a relatively large body of literature has shown that the unrestricted EFA model is often better suited to reflect the complex factorial structure of Big Five scales than the more restrictive "simple structure" CFA approach (e.g. [Asparouhov & Muthén, 2009](#); [Booth & Hughes, in press](#); [Borkenau & Ostendorf, 1990](#); [Hopwood & Donnellan, 2010](#); [Marsh et al., 2010](#)). This is mainly because CFA, or more precisely, its notion of independent clusters of items assumes that indicators perfectly measure only one target trait, though this condition is rarely met.

Second, the literature shows that self-report measures are commonly plagued by different sorts of method bias (see, for an overview, [Podsakoff, MacKenzie, & Podsakoff, 2012](#)). Particularly, acquiescence response style, also known as directional or uniform response bias (hereafter ARS), has been identified as one of the most important nuisance factors in personality measurement (e.g. [McCrae, Herbst, & Costa, 2001](#); [Rammstedt & Farmer, 2013](#)). It therefore has become standard to mitigate this problem by using

semantically balanced scales in questionnaires, i.e., using positively (pro-trait) and negatively (con-trait) worded items.

Most importantly, ARS can distort the factorial structure, i.e., convergent and discriminant validity of the questionnaire items as systematic measurement error is introduced ([Podsakoff et al., 2012](#)). As a consequence, the "true" factorial structure, especially in multidimensional instruments (e.g. Big Five scales), is misrepresented. Several approaches have therefore tried to remedy this problem by means of adjustment techniques that mitigate the impact of ARS, for instance, by using a CFA model of ARS ([Billiet & McClendon, 2000](#); [Maydeu-Olivares & Coffman, 2006](#)) or by subtracting the within-person mean response from each item, also called ipsatization ([Rammstedt & Farmer, 2013](#)).

## 2. Aim and contribution of this study

In this study I investigate the properties of Random Intercept EFA (RI-EFA) ([Aichholzer, submitted for publication](#)). RI-EFA treats ARS as an individual random intercept (RI) factor extending the standard EFA model. In particular, the RI-EFA approach is applicable whenever multidimensional and semantically balanced scales of psychological constructs or attitudes are examined. This research demonstrates the usefulness of RI-EFA using two standard Big Five personality scales, the German versions of the NEO-FFI ([Borkenau & Ostendorf, 1993](#)) and the BFI ([Rammstedt & John, 2005](#)). In addition, I test four hypotheses that have been put forward in previous research:

- H<sub>1</sub>: Measurement models with complex item-factor structure (EFA) commonly fit better to Big Five data than restricted simple structure models (CFA) (e.g. [Hopwood & Donnellan, 2010](#); [Marsh et al., 2010](#)).

\* Fax: +43 1 4277 9499.

E-mail address: [julian.aichholzer@univie.ac.at](mailto:julian.aichholzer@univie.ac.at)

- H<sub>2</sub>: Measurement models with adjustment for ARS fit the data better than models without such adjustment (e.g. Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006).
- H<sub>3</sub>: ARS adjustment improves the clarity of factor loading patterns in ways that approach the ideal “perfect simple structure” form (e.g. McCrae et al., 2001; Rammstedt & Farmer, 2013).
- H<sub>4</sub>: ARS adjustment improves the clarity of factor loading patterns in particular in populations that are prone to ARS (e.g. Rammstedt & Farmer, 2013; Rammstedt, Goldberg, & Borg, 2010).

H<sub>1</sub> is tested by comparing common model fit criteria of standard CFA with EFA. H<sub>2</sub> is tested by comparing RI-EFA with the other models. H<sub>3</sub> and H<sub>4</sub> are assessed by comparing empirical factor loading structures with a perfect simple structure matrix (i.e., perfect  $-1/0/+1$  entries). The contribution of this study is thus twofold: First, a novel method in the field of factor analysis is tested with regard to its applicability to established scales. Second, these scales are re-evaluated using this new method.

### 3. The RI-EFA model

Recently, Aichholzer (submitted for publication) has presented the RI-EFA model that allows researchers to combine standard EFA, i.e., an unrestricted or complex item-factor loading matrix, with a random intercept (RI) factor that represents ARS or uniform response bias more generally. So far, the latter option was only available for restricted CFA models (Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006).

The RI-EFA model is an extension of the common factor model and reads as follows

$$Y_{ki} = \tau_k + \Lambda^* \cdot \eta_{ji} + \mathbf{1} \cdot \alpha_i + \varepsilon_{ki}$$

where the  $k \times 1$  vector of observed scores  $Y_{ki}$  for a respondent  $i$  is derived from item intercepts  $\tau_k$ , unconstrained loadings  $\lambda_{kj}$  on a hypothesized number of  $j$  latent factor scores  $\eta_{ji}$  that are contained in the  $k \times j$  factor loading matrix  $\Lambda$ , the individual RI/ARS factor  $\alpha_i$ , and a vector of all unique (residual) factors  $\varepsilon_{ki}$ . Note that the factor loading matrix  $\Lambda$  has to be rotated by some criterion (i.e., by orthogonal or oblique rotation) for the model to be identified ( $\Lambda'$  after rotation). Common assumptions in factor analysis apply, namely that factors are uncorrelated with residuals and residuals are mutually uncorrelated. Regardless of the rotation method used for the latent “substantial” factors,  $\alpha$  is assumed to be orthogonal to factors and item uniquenesses for identification, i.e.:

$$\begin{aligned} \text{Cov}(\eta_j, \varepsilon_k) &= \text{Cov}(\varepsilon_k, \varepsilon_{k'}) = 0 \text{ for } \varepsilon_k \neq \varepsilon_{k'} \text{ and } \text{Cov}(\eta_j, \alpha) \\ &= \text{Cov}(\varepsilon_k, \alpha) = 0 \end{aligned}$$

Further note that the factor  $\alpha$  has a constant loading vector of  $\mathbf{1}$  on all indicators, regardless of their keying (pro-trait or con-trait), while its variance is freely estimated and tested to be non-zero.<sup>1</sup> The RI-EFA model is thus a hybrid model that combines an EFA part where item-factor loadings are freely estimated and a restricted CFA part where item-factor loadings on the RI/ARS factor  $\alpha$  are restricted to follow a predefined pattern. Due to its specific factor loading structure  $\alpha$  must not be confused with other factors of personality that load on all items (see Anusic, Schimmack, Pinkus, & Lockwood, 2009). In other words,  $\alpha_i$  represents a uniform shift of individual item responses independent from substantial factors. Hence, it can be considered an issue of differential item functioning (DIF) or violation of measurement invariance.

Furthermore, the model can be estimated quite easily applying the Exploratory Structural Equation Modeling (ESEM) framework (Asparouhov & Muthén, 2009) and the Mplus software (Muthén & Muthén, 1998–2012) (see the Appendix A for example Mplus code).

## 4. Materials and methods

### 4.1. Instruments

This study investigates the German 60-item NEO-FFI (Borkenau & Ostendorf, 1993) and the German 44-item BFI (Rammstedt & John, 2005). Both the NEO-FFI and the BFI are based on the Big Five taxonomy of Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to experience (O). All items were measured on a 5-point scale with endpoints labeled as 1 – *strongly disagree* to 5 – *strongly agree* in the NEO-FFI and 1 – *does not apply at all* to 5 – *applies completely* for self-reports in the BFI. In the sample used here, the Cronbach's Alpha estimates for the hypothesized Big Five dimensions were .75 (E), .72 (A), .84 (C), .82 (N), and .66 (O) for the NEO-FFI and .81 (E), .76 (A), .80 (C), .74 (N), and .82 (O) for the BFI. Note that previous research has already investigated the factorial structure of these instruments using different factor analytic strategies (Booth & Hughes, in press; Marsh et al., 2010; Rammstedt & Farmer, 2013).

### 4.2. Data

The NEO-FFI and the BFI were administered in a random sample of the German population (aged 18 and above) as part of a larger study on personality and political behavior (Schumann, 2004).<sup>2</sup> The total sample with valid demographic data comprised  $n = 2508$  respondents. Age ranged between 18 and 92 ( $Mean = 49$ ,  $SD = 17$ ) and 52% of the respondents were female, 48% male. The NEO-FFI was administered face-to-face, while the BFI was administered using self-administration as a drop-off (with  $n = 1492$  participants in total).

### 4.3. Analysis

Analyses were conducted with standard simple structure CFA, EFA (or simple ESEM), and RI-EFA using linear MLR (maximum-likelihood with robust standard errors) estimation as well as WLSMV (weighted least square mean- and variance-adjusted) estimation for ordered categorical measures in Mplus 7 (Muthén & Muthén, 1998–2012). Global goodness-of-fit indices were inspected for each model in addition to the  $\chi^2$ -test. The reason for this is that  $\chi^2$ -tests commonly result in a rejection of the model when applied to large samples. According to common fit criteria (see Marsh, Hau, & Wen, 2004), acceptable fit is achieved when CFI > .90, TLI > .90, and RMSEA < .08, whereas excellent fit is achieved when CFI > .95, TLI > .95, and RMSEA < .05. Further, better fit of a model is also supported by lower BIC values.

For all empirical analyses the theoretical five-factor structure was imposed to investigate the hypothesized structure of the NEO-FFI and the BFI. Solutions for the rotated factor loading matrix in the EFA part were computed applying oblique Quartimin rotation (for other factor rotation criteria see Asparouhov & Muthén, 2009).

## 5. Results

I start by comparing the fit measures for the different modeling strategies (Table 1). First, the results reconfirm previous evidence suggesting that unrestricted EFA (or simple ESEM) fits Big Five data

<sup>1</sup> Note that other approaches also follow the assumption of a uniform bias, for instance, when a constant is subtracted from all individual item responses (ipsatization).

<sup>2</sup> Note that the item ordering differs from the original NEO-FFI. Item\_1 and Item\_21 were interchanged.

**Table 1**  
Summary of goodness-of-fit indices for different modeling approaches.

Measure	Method	Model	$\chi^2$	d.f.	$p(\chi^2)$	CFI	TLI	RMSEA	BIC
NEO-FFI ( $n = 2411$ )	MLR	CFA	12,526	1,700	<.01	.661	.647	.051	392,714
		EFA	4,716	1,480	<.01	.899	.879	.030	385,128
		RI-EFA	4,273	1,479	<.01	.913	.895	.028	384,607
	WLSMV	CFA	24,644	1,700	<.01	.683	.670	.075	n.a.
		EFA	5,907	1,480	<.01	.939	.927	.035	n.a.
		RI-EFA	5,078	1,479	<.01	.950	.940	.032	n.a.
BFI ( $n = 1428$ )	MLR	CFA	7,074	892	<.01	.640	.618	.070	165,411
		EFA	2,404	736	<.01	.903	.875	.040	160,651
		RI-EFA	1,787	735	<.01	.939	.921	.032	159,896
	WLSMV	CFA	13,807	892	<.01	.656	.635	.101	n.a.
		EFA	3,470	736	<.01	.927	.906	.051	n.a.
		RI-EFA	2,302	735	<.01	.958	.946	.039	n.a.

Note: A five-factor solution was hypothesized in all cases. EFA solutions use Quartimin rotation. All models converged. n.a. = not available. No post-stratification weights were applied. Sample size  $n$  refers to listwise deletion.

considerably better than standard simple structure CFA (see Table 1), which supports  $H_1$ .

Second, having established that EFA fits better than CFA, it is shown that RI-EFA has a better model-to-data fit than EFA for the NEO-FFI and the BFI, which supports  $H_2$ . Better fit for RI-EFA vs. EFA is further supported by lower BIC values, significant  $\chi^2$  reduction (each  $p < .01$  at  $-1$  degrees of freedom), and rejection of  $Var(\alpha) = 0$  (at  $p < .01$ ). If the RI/ARS bias was irrelevant, the model fit should be very similar or equal. Note that these findings hold regardless of the estimation method (MLR or WLSMV).

Third, as suggested by Hopwood and Donnellan (2010), I compared the congruence of the empirical factor loading pattern to a perfect simple structure matrix by assigning each item to its hypothesized factor, using Tucker’s congruence coefficient  $c$  (see Table 2, detailed factor loading results using MLR estimation are available as supplemental materials). After partialling out the additional RI/ARS factor by means of RI-EFA, there is convincing evidence in support of further alignment ( $\Delta$  of congruence) of items towards the hypothesized five-factor personality structure, which confirms  $H_3$ . In other words, the matrix congruence to perfect simple structure is always larger for the RI-EFA factor loading patterns. Further, the variance explained by the RI/ARS factor can be computed using squared standardized loadings, 2.8% (NEO-FFI) and 7.5% (BFI) with MLR estimation, which is largely similar to previous findings (3–4%) (e.g. Anusic et al., 2009; Billiet & McClelland, 2000).

Fourth, I look at subpopulations with potentially different amounts or unequal variance in ARS. A common proxy is the respondent’s educational level (Rammstedt et al., 2010) which was operationalized by contrasting “low” education (lower secondary education or less) vs. “high” education (admission to tertiary education or completed university degree), whereas “medium” education was omitted. The results reconfirm that the factor loading congruence to a hypothesized five-factor simple structure is, in general, lower among less educated respondents.

**Table 2**  
Estimates of factor loading matrix congruence to perfect simple structure.

Measure	Model	Total sample	Low/high education
NEO-FFI	EFA	.78	.73/.77
	RI-EFA	.84	.78/.83
	$\Delta$	.06	.05/.06
	$n$	2411	1095/491
	BFI	EFA	.73
	RI-EFA	.90	.86/.90
	$\Delta$	.17	.15/.09
	$n$	1428	677/290

Note: A five-factor solution was hypothesized in all cases. Entries indicate congruence coefficient  $c$  for the MLR Quartimin rotated solution. Sample size  $n$  refers to listwise deletion.

Applying RI-EFA improves the clarity of factor loading patterns among both groups, though it improves especially among less educated respondents who filled out the BFI, which partly supports  $H_4$ .

In sum, the results suggest that applying RI-EFA makes a greater difference for revealing the “true” structure in the case of the BFI, whereas the impact of response bias on substantial results seems to be smaller for the NEO-FFI. However, it should be mentioned that differences between the two instruments may also reflect the use of different survey modes (F2F vs. P&P).

## 6. Discussion

This research shows that RI-EFA can be used as an effective factor-analytic method. It provides a useful tool for researchers when they would like to investigate the factorial structure of multidimensional scales based on semantically balanced or heterogeneous items. The results support the improvement of model-to-data fit and suggest that the hypothesized factor structure became “clearer” (i.e., it approximated simple structure). Moreover, given that the identified bias resembles ARS, it may be the preferred method whenever heterogeneous populations with potentially different response behavior are studied or when specific measurement conditions apply that evoke such response bias.

The main advantage which makes RI-EFA superior to other approaches, such as Principal Component Analysis (PCA) of ipsatized data (Rammstedt & Farmer, 2013), is that RI-EFA fits into the larger family of common factor models and structural equation modeling or ESEM (Aichholzer, submitted for publication). Therefore RI-EFA can be extended to testing measurement invariance over subgroups or over time as well as to testing covariates of the RI/ARS factor and, hence, causes of such bias. As has been shown, RI-EFA can also be extended to the case of ordered categorical indicators but also to binary items using WLSMV estimation (Asparouhov & Muthén, 2009).

Nevertheless, there are some limitations to RI-EFA, some of which apply to the ESEM framework more generally (e.g. Booth & Hughes, in press; Herrmann & Pfister, 2013; Marsh et al., 2010). Firstly, RI-EFA is neither a strictly exploratory nor a strictly confirmatory method, but it might be valuable as a “first-step” method in the sense that it is less rigorous compared to the CFA approach. Second, different rotation criteria (Quartimin, Geomin, Varimax, etc.) should be examined as factor loadings and between-trait correlations are dependent on the rotation criterion used. Third, it is still a matter of debate to what extent the factor scores and associations with covariates will differ between the different approaches, i.e., CFA, EFA/ESEM, or RI-EFA (also see Booth & Hughes, in press; Herrmann & Pfister, 2013). Apart from that, quite often the items in a scale are not fully but partially balanced (e.g. NEO-FFI and BFI), that is, some subscales use an even number of

pro-trait and con-trait items, whereas others are primarily keyed in one direction. Indeed, the ultimate goal of the RI/ARS factor is that it conveys all information from balanced subsets of items to also control for ARS in other items and subscales that are partly or completely unbalanced. One could also let a selection of balanced item sets define the RI/ARS factor (for a set of items in the BFI see Rammstedt & Farmer, 2013), whereas loadings are set to +1, and freely estimate loadings of the remaining items on that factor. Such a model can then be compared to the basic RI-EFA model (constant loadings of +1). Still, as previously noted (Aichholzer, submitted for publication), the construct validity of the RI/ARS factor itself will be higher with an increasing number of balanced items that come from heterogeneous constructs. However, less is known about the amount of balancing in scales in order to fully identify and accurately address the type of bias which is modeled here. Future studies might investigate these issues in more depth.

### Acknowledgments

This research is conducted under the auspices of the Austrian National Election Study (AUTNES), a National Research Network (NFN) sponsored by the Austrian Science Fund (FWF) [S10902-G11]. I also would like to thank Kathrin Thomas, Markus Wagner, and the two anonymous referees for helpful comments on earlier versions of this paper.

### Appendix A. Example Mplus code for the RI-EFA model

---

```
Variable:
  NAMES = item1-item60; ! Example for 60-item NEO-
  FFI
  !CATEGORICAL = item1-item60; ! optional
Analysis:
  ESTIMATOR = MLR; ! Choose estimator: MLR, WLSMV,
  etc.
  ROTATION = QUARTIMIN; ! Select rotation criterion
Model:
  F1-F5 BY item1-item60 (*L); ! Define 5-factor
  Lambda matrix
  RI BY item1-item60@1; ! Define RI/ARS factor
  loadings
  RI WITH F1-F5@0; ! RI/ARS is orthogonal
Output:
  STDYX; ! Compute fully standardized factor
  loadings
```

---

### Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jrp.2014.07.001>.

### References

- Aichholzer Julian (submitted for publication). Testing complex factor structure and individual method bias with random intercept EFA. *Multivariate Behavioral Research*.
- Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology*, 97(6), 1142–1156.
- Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438.
- Billiet, J. B., & McClelland, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7(4), 608–628.
- Booth, T., & Hughes, D. J. (in press). Exploratory structural equation modeling of personality data. *Assessment*. <http://dx.doi.org/10.1177/1073191114528029>.
- Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, 11(5), 515–524.
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Herrmann, A., & Pfister, H.-R. (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five inventories? *Journal of Research in Personality*, 47(5), 599–608.
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14(3), 332–346.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341.
- Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., & Trautwein, U. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471–491.
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362.
- McCrae, R. R., Herbst, J. H., & Costa, P. T. Jr. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Berlin: Pabst.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7 ed.). Los Angeles, CA: Muthén & Muthén.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569.
- Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, 25(4), 1137–1145.
- Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44(1), 53–61.
- Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K). *Diagnostica*, 51(4), 195–206.
- Schumann, S. (2004). *Personality and voting behaviour 2003*. Cologne: GESIS Data Archive (ZA4052 Data file Version 1.0.0).