Journal of Research in Personality 47 (2013) 599–608

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five inventories?

Anne Herrmann [a,*], Hans-Rüdiger Pfister [b]

[a] Kalaidos University of Applied Sciences Switzerland, Department of Economics and Management, Jungholzstrasse 43, 8050 Zurich, Switzerland
[b] Leuphana University of Lueneburg, Institute of Experimental Industrial Psychology, Wilschenbrucher Weg 84a, 21335 Lueneburg, Germany

ARTICLE INFO

Article history:
Available online 18 May 2013

Keywords:
Personality
Big Five structure
Confirmatory factor analysis
Exploratory Structural Equation Modeling
Construct validity
Multitrait–multimethod
NEO PI-R
16PF

ABSTRACT

The poor performance of five-factor personality inventories in confirmatory factor analyses (CFAs) prompted some to question their construct validity. Others doubted the CFA's suitability and suggested applying Exploratory Structural Equation Modeling (ESEM). The question arises as to what impact the application of either method has on the construct validity of personality inventories. We addressed this question by applying ESEM and CFA to construct better-fitting, though more complex models based on data from two questionnaires (NEO PI-R and 16PF). Generally, scores derived from either method did not differ substantially. When applying ESEM, convergent validity declined but discriminant validity improved. When applying CFA, convergent and discriminant validity decreased. We conclude that using current personality questionnaires that utilize a simple structure is appropriate.

1. Introduction

Researchers who investigate normal adult personality have reached a consensus on five broad factors, often called the 'Big Five' (Goldberg, 1990), and on their conceptual definitions (Digman, 1990; McCrae & Costa, 1999; Norman, 1963). These factors are known as Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, although other terms are used as well. This general consensus has allowed for cumulative research and meta-analyses of important aspects of the construct, including the development of personality over an individual's lifespan (Judge, Higgins, Thoresen, & Barrick, 1999; Terracciano, McCrae, & Costa, 2010), differences between groups (Goldberg, Sweeney, Merenda, & Hughes, 1998; Schmitt, Realo, Voracek, & Allik, 2008), the existence of a general factor of personality (Musek, 2007; van der Linden, te Nijenhuis, & Bakker, 2010), a prediction of external criteria (Grucza & Goldberg, 2007; Hurtz & Donovan, 2000), and many more. In research and practice, personality is predominantly assessed using self-report questionnaires. Many of these questionnaires contain items that contribute to one of many first-order scales that are combined to represent the Big Five factors.

The internal structure of personality, i.e., the assignment of subscales to the five factors, has commonly been examined using an exploratory factor analysis (EFA; Aluja, Rossier, Garcia, & Verardi, 2005; Cattell & Cattell, 1995; Costa & McCrae, 1992b). This assignment is extremely important because it forms the basis for obtaining scores for the higher-order personality factors. In general, a simple structure (Thurstone, 1947) where each first-order scale is uniquely assigned to only one of the Big Five factors is assumed to be appropriate.

As in many other research areas in which constructs are assessed using self-report questionnaires, CFAs were eventually applied to personality data. The results of these studies were largely discouraging. The CFA model fit indices frequently exceeded proposed cut-off values for acceptable model fits and, based on CFA standards, did not confirm the simple structure (Church & Burke, 1994; Hopwood & Donnellan, 2010; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Vassend & Skrondal, 2011). Several cross loadings (i.e., links between first-order scales and factors other than the originally postulated higher-order personality factors) usually needed to be included in the model to achieve an acceptable fit. The more complex models, however, were difficult to interpret and often displayed less of a good fit in cross-validation samples (e.g., Church & Burke, 1994; Hopwood & Donnellan, 2010).

This has raised concerns if the currently proposed composition of the broad factors provides an adequate assessment of an individual's personality. These higher-order scores are commonly used in research studies and in practical applications of personality instruments. Thus, confidence is required regarding the suitability of the Big Five factors as a 'common language' for describing personality. Adding additional cross loadings as suggested by CFA also changes the meaning of the observed scores. Subsequently, one must

* Corresponding author.
E-mail address: anne.herrmann@kalaidos-fh.ch (A. Herrmann).

0092-6566/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jrp.2013.05.004

question how the construct validity of personality instruments is affected when subscales contribute to more than one broad factor.

In the present study we address these concerns in two ways: First, we determine the 'change of scores' which – in this examination – refers to a difference in the relative position of an individual within a sample on the trait continuum measured as the correlation between the original scores and scores obtained after incorporating the CFA cross loadings. Second, we examine the impact on the instruments' construct validity resulting from the modified models.

To complement our investigation and consider more recent trends in factor analysis, we also apply Exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthen, 2009), a method that integrates CFA and EFA. ESEM is less restrictive than CFA as it does not constrain the non-target loadings to be zero. In difference to CFA, in ESEM a model can be specified only with regard to the number of factors. Further restrictions can be added and tested using chi-square difference tests. In difference to EFA, ESEM provides typical CFA parameters, such as standard errors and goodness of fit statistics as well as the possibility to test for measurement invariance between groups and across time (Asparouhov & Muthen, 2009). Due to these possibilities and advantages of ESEM, it has been promoted to be applied in the psychometric evaluation of psychological instruments (Marsh, Liem, Martin, Morin, & Nagengast, 2011).

We applied a CFA and ESEM to data from 620 respondents who completed two established personality questionnaires (the NEO PI-R and the 16PF questionnaire). Using two different sets of modification criteria to determine cross loadings when conducting the CFA, we generated two more complex models for each instrument. We computed scores based on these modified CFA models using two different approaches: (a) we applied the scoring rules for the instrument provided in the respective test manual but added the additional subscales, as identified in the CFA and (b) we used the factor scores obtained from the respective modified CFA model. The first approach mirrors current usage in research, in which manifest, rather than latent, Big Five scores are employed (Barrick & Mount, 1996; Grucza & Goldberg, 2007; Hurtz & Donovan, 2000; Salgado, 2003). The second approach uses scores that correspond more directly with the CFA models. With regard to the application of ESEM, we used the factor scores obtained from applying the method from both instruments.

To assess the relative score changes, we computed correlations between scores from the original model and the scores obtained from the CFA and ESEM models. The results of this analysis support a more nuanced discussion of the discrepancy between current personality theories and the more complex model of personality, as suggested by the CFA. Applying ESEM offers further insight into how Big Five scores based on a more recent factor-analytical method.

To determine the impact on the questionnaires' construct validity, we applied the multitrait–multimethod (MTMM) approach, which was developed by Campbell and Fiske (1959), to the original model as well as the models proposed by CFA and ESEM. A comparison of the MTMM results across the models showed the extent to which the relationships within and between the five factors of both instruments changed as one moved from a simple to a more complex structure, thus determining changes in the convergent and discriminant validity.

Previous studies have focused mainly on investigating the congruence between results obtained from the EFA and CFA of an instrument without examining the impact of the observed discrepancies on scale scores and construct validity beyond the internal structure (e.g., Aluja, Blanch, & Garcia, 2005; Borkenau & Ostendorf, 1990; McCrae et al., 1996). In other studies, CFAs were applied to several instruments, but it was not determined how the relationships between the constructs were affected by changes in the model proposed by the CFAs (e.g., Church & Burke, 1994; Hopwood & Donnellan, 2010). In our study, we address those gaps by determining how the scores of and the relationships between personality scales change when the internal structure is more complex, as suggested by CFA. As a result, we extend the examination of construct validity beyond the internal structure to focus on changes in the convergent and discriminant validity within and across the two instruments. The study thus follows a suggestion made, among others, by Hopwood and Donnellan (2010) that "there is a need to document that misspecifications have practical or substantive consequences beyond simply contributing to model misfit" (p. 343).

Considering the complexities and difficulties in identifying the correct model in CFA based on modification indices and other model assessment criteria (Fan & Sivo, 2007; MacCallum, Roznowski, & Necowitz, 1992), we do not aim at determining the "true" model of personality. Instead, we provide an empirical illustration, i.e., to demonstrate by way of example the impact that this added complexity would have on scores and construct validity. By also applying ESEM to both instruments, we shed light on how this more recent but increasingly used method may affect the resulting factor scores and subsequently the instruments' construct validity.

2. Method

2.1. Measures

The data from two hierarchical self-report personality instruments were used in this study:

(1) Cattell's 16 Personality Factor Questionnaire, 5th Edition (16PF, Conn & Rieke, 1994) consists of 185 items with a three-choice response format that measures 16 primary factors. The 15 non-cognitive factors are then combined into five factors, commonly called 'global factors'.
(2) The Revised NEO Personality Inventory (NEO-PI-R, Costa & McCrae, 1992b) comprises 240 items with a five-point Likert response format. It assesses 30 facets of personality that are used to compute five higher-level domain scores.

The 16PF and the NEO-PI-R differ in that the first-order level of personality is described with 15 and 30 scales, respectively. An alignment exists, however, between the second-order level, where there is a NEO domain counterpart for each 16PF global factor. The counterparts for both instruments are 16PF-Extraversion and NEO-Extraversion, 16PF-Anxiety and NEO-Neuroticism, 16PF-Self-Control and NEO-Conscientiousness, 16PF-Independence and NEO-Agreeableness and, finally, 16PF-Tough-Mindedness and NEO-Openness (Cattell & Mead, 2008). The last two pairs are defined in the opposite direction.

Different views exist on when to consider a psychometric questionnaire a "Big Five Instrument". We follow a definition by McCrae and John (1992): "The five-factor model of personality is a hierarchical organization of personality traits in terms of five basic dimensions" (p. 175) which applies to the NEO PI-R as well as the 16PF. These two Big Five instruments were included in this study because they differ profoundly in their development and in the approach to computing the second-order factors. This method safeguards against drawing conclusions about personality constructs that are actually a result of characteristics of a particular instrument. The 16PF questionnaire was developed based on empirical analyses. An EFA of the item parcels was carried out to identify the primary personality traits. These primary factors were subjected to a second-order EFA to extract five global factors (Cattell