Personality and Individual Differences 50 (2011) 1300–1304

Contents lists available at ScienceDirect

Personality and Individual Differences

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

The NEO personality inventory revised (NEO-PI-R): Exploring the measurement structure and variants of the five-factor model

Olav Vassend [a,*], Anders Skrondal [b,c]

[a] Department of Psychology, University of Oslo, Norway
[b] Norwegian Institute of Public Health, Oslo, Norway
[c] London School of Hygiene and Tropical Medicine, London, UK

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form 25 February 2011
Accepted 1 March 2011

Keywords:
NEO-PI-R
The five-factor model
Measurement models
Confirmatory factor analysis

ABSTRACT

Despite the popularity of the NEO-PI-R as a measure of the five-factor model (FFM), several questions regarding its psychometric foundation have yet to be investigated. Using NEO-PI-R data obtained in a US community sample, confirmatory factor analyses demonstrated that most of the single-factor models of individual facets and domains evidenced acceptable fit. However, one-factor models of the domains Extraversion and Agreeableness were not supported, which calls into question their adequacy as measurement models. Analyses of variants of the FFM revealed that only the unrestricted exploratory factor model showed acceptable fit as well as replicability across gender and educational level.

1. Introduction

Over the past two or three decades, the five-factor model (FFM) of personality has become one of the dominant paradigms in trait psychology (McCrae & Allik, 2002). One of the most popular measures of the FFM is the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992). Based on extensive cross-cultural studies using translations of this instrument, McCrae, Costa, and coworkers (e.g., McCrae & Allik, 2002; McCrae & Terracciano, 2005; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996) maintain that the FFM is highly uniform across most cultures and that it probably reflects a universal structure of individual differences in personality.

Despite this apparent consensus among the FFM proponents, several concerns have been raised regarding the assessment and interpretation of model structure and content (e.g., Block, 1995; Eysenck, 1992; Gignac, Bates, & Jang, 2007). However, criticisms of the FFM have chiefly focused on structural aspects of the model (e.g., factor analytic studies of how NEO-PI-R facets and domains are related), whereas the basic measurement model specifying the relationship between the latent variables and the observed variables (items), is taken more or less for granted. Thus, studies using the NEO-PI-R generally employ scoring criteria derived from the original test manual (Costa & McCrae, 1992). A second central issue involves the rather uncritical adoption and widespread use of exploratory factor analysis (EFA) in the FFM tradition. It is worth noting that satisfactory replications across languages and cultures are generally achieved only through EFA-based methods and criteria, particularly the use of the orthogonal Procrustes rotation method and analysis of factor pattern congruence (e.g., McCrae et al., 1996).

Unlike results from the EFA-based research on the FFM, in studies using confirmatory factor analysis (CFA) structural inconsistencies as well as instances of conspicuous model misfit have been demonstrated repeatedly (e.g., Church & Burke, 1994; McCrae et al., 1996). Yet, in a more recent study Aluja and collaborators (Aluja, García, García, & Seisdedos, 2005) claim that a five-factor structure of the NEO-PI-R facets are supported as judged by the usual EFA procedures as well as CFA-based goodness of fit indices. However, only models in which very small loadings were freely estimated were judged to be satisfactory and all goodness of fit indices were not unequivocally within the accepted limits.

Somewhat paradoxically, however, the discrepant EFA and CFA findings have lead several proponents of the FFM to question the utility of CFA in studies of complex personality structure (Hopwood & Donnellan, 2010). McCrae et al. (1996) have gone so far as to claim that CFA is systematically flawed, meaning that it can reject empirically sound models and accept models that are not. Yet, while McCrae et al. (1996) are highly critical of the use of the likelihood ratio test and goodness of fit indices in CFA as a means of assessing model fit, they acknowledge that some formal evaluation of factor models and factor replicability is needed. However, although their preferred method, congruence analysis, may

* Corresponding author. Address: Department of Psychology, University of Oslo, P.O. Box 1094, Blindern, NO-0317 Oslo, Norway. Tel.: +47 22845197.
E-mail address: olav.vassend@psykologi.uio.no (O. Vassend).

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.03.002

be appropriate as a preliminary step to aid in the formulation of suitable hypotheses (McDonald, 1985) and may also be sufficiently accurate to examine factorial similarity at a global level, it is inadequate when one wants to identify more subtle structural characteristics (parameters) of particular models or structural differences across groups. Thus, essentially bad-fitting models, such as orthogonal NEO-PI-R models (as suggested by e.g., McCrae et al., 1996) can be judged replicable using congruence coefficients. Furthermore, as explained by Millsap and Meredith (2007), descriptive indices of congruence among rotated factor patterns are not fit of measures to the data; they only assess similarities of factor structure across groups. There is now general agreement that CFA and multi-group CFA (Jöreskog, 1971) represents the most powerful and versatile approach for the assessment of various aspects of measurement models, including factorial invariance (Millsap & Meredith, 2007).

Several researchers have raised the issue whether even extensively modified versions of the FFM can be viewed as valid representations of the 30 facets within the NEO-PI-R (e.g., Gignac et al., 2007). Gignac et al. (2007) maintain that the main problem has to do with the instrument's comprehensiveness and as yet poorly understood complexity, not any technical limitation with CFA as such (for similar views, see Vassend & Skrondal, 1997). As a consequence, they recommend personality researchers to consider developing measures with a more restricted scope than is usually the case within the FFM tradition. In a recent paper, Gignac (2009) notes that assessment of single-factor models of each personality dimension within the FFM framework does not appear to have been attempted or suggested in the past.

To an increasing extent, personality constructs are included in comprehensive research projects in areas such as psychopathology (Krueger & Tackett, 2006) or health psychology (Vollrath, 2006). As is often the case, advanced statistical techniques like structural equation modeling (SEM) are being used. As noted by Brown (2006), when poor model fit is encountered in SEM research it is more likely that it stems from misspecifications in the measurement (factor) model than from the structural portion of the model that specifies the relationships among the latent variables. Essentially, if measurement models do not fit the observed variables, then relationships among the latent variables in structural models are rendered ambiguous or at worst meaningless. Thus, the issue of factor model specification and assessment should be viewed in the larger context of the enterprise of searching for acceptable measurement models in the personality field.

In the present study, a factor analytic strategy based mainly on CFA was developed in order to expose structural complexities in the NEO-PI-R instrument in a systematic and perspicuous way. Thus, several factor models representing both the original as well as alternative conceptions of the instrument's latent structure were assessed and compared. Following Gignac et al. (2007), structural integrity of each of the five domains and the 30 facets was examined. Moreover, factorial invariance of plausible factor models across gender and educational levels was investigated. Consequences of the research findings for the interpretation of structural properties of the NEO-PI-R are discussed, and suggestions for model re-specifications within the FFM framework are presented.

2. Methods

2.1. Participants

The sample included 478 women and 378 men in the age range 18–85 (Mean = 50.8, *SD* = 13.2), recruited from the Eugene-Springfield (Oregon, USA) community (see Grusza & Goldberg, 2007). The participants had agreed to work with the Oregon Research Institute personality team for at least five years, and they were periodically mailed questionnaires and inventories, which they were paid to complete. Education level of the subjects was rated on an 8-point scale (Mean = 5.77, *SD* = 1.65).

2.2. Personality measure

The NEO-PI-R was designed explicitly according to the FFM. It is constructed as a three-level instrument, comprising items (240 plus a validity question), facets (30), and domains (the five broad dimensions, commonly referred to as Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness; Costa & McCrae, 1992).

2.3. Factor analytic strategy

The latent structure of the NEO-PI-R was explored at both the facet and item level in the factor analytic study, which was conducted in several steps.

1. Based on results from previous factor analytic research (McCrae et al., 1996), several variants of the FFM were assessed. First, the original *simple structure orthogonal model* (Model 1a) was examined. In this model 25 free parameters were estimated, while five were set to unity (one facet for each of the five factors). The remaining 120 fixed loadings were set to zero. Based on criteria suggested by McCrae et al. (1996), two re-specified orthogonal five-factor models were then considered (Models 1b and 1c). However, instead of fixing secondary factor loadings to values obtained in normative samples, in the present analyses these parameters were freely estimated. In a *salient-loadings model* (Model 1b) parameters were freed according to the factor loadings estimated by EFA (i.e., loadings greater than ±.40). The other fixed loadings were set to zero. In a *modest-loadings model* (Model 1c) EFA-estimated loadings greater than ±.20 were freely estimated; the other loadings were set to zero. Models 1a-1c were then re-estimated as Models 2a-2c by relaxing the constraint that the factors be orthogonal.
2. The procedure of "exploratory factor analysis within the CFA framework" (E/CFA; Jöreskog, 1969) was then conducted. In this analysis, the CFA applies the same number of identifying restrictions used in EFA and produces the same model fit as maximum likelihood EFA. However, the CFA estimation provides additional important information, such as statistical significance of factor loadings and factor covariances.
3. Factorial invariance across gender and educational level of models judged acceptable in the preceding analyses, was investigated in multi-group CFA (Jöreskog, 1971). The sample was divided into a low-education group (i.e., subjects with education level 1–5 [5 = some college education], $n = 393$) and a high-education group (i.e., subjects with education level 6–8 [8 = post-college degree], $n = 454$). In such analyses, only the factor pattern matrices should be treated as possibly invariant across sub-populations (Vassend & Skrondal, 1997). Hence, the factor covariances and unique variances were estimated independently in the groups, whereas the factor loadings were estimated under the assumption that they are equal for all groups.
4. In the next series of analyses, each of the five dimensions of the NEO-PI-R was modeled, individually, as one-factor models. If the posited one-factor model was found acceptable, factorial invariance analysis was conducted as detailed above.
5. Finally, structural integrity at the facet level was investigated. Thus, each of the 30 facets was modeled as one-factor models.