Personality and Individual Differences 52 (2012) 503–508

Contents lists available at SciVerse ScienceDirect

Personality and Individual Differences

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

The Five Factor Model in personnel selection: Measurement equivalence between applicant and non-applicant groups

Michele Vecchione *, Guido Alessandri, Claudio Barbaranelli

"Sapienza" University of Rome, 00185 Rome, Italy

ARTICLE INFO

Article history:
Received 15 August 2011
Received in revised form 16 November 2011
Accepted 21 November 2011
Available online 16 December 2011

Keywords:
Measurement invariance
Applicants
Five Factor Model
Personnel selection
Applied settings

ABSTRACT

This study investigates the measurement equivalence of a Five Factor measure of personality between applicant and non-applicant samples. The Big Five Questionnaire-2 was administered in two samples: A group of volunteers ($n = 903$), who completed the test for research purposes, and a group applying for jobs, who completed the test during hiring procedures ($n = 401$). Multiple-group confirmatory factor analysis on item composites was conducted to test for the equivalence of factor covariance and mean structures of twenty facets of the Five Factors. Item-level analyses were carried out through analysis of variance to further examine the issue of measurement invariance. Findings suggested that personality facets have the same measurement unit across applicants and non-applicants, while a lack of equivalence was found in the origin of the scales. Similar results were found at the item-level. Implications for personality assessment are advanced and discussed.

1. Introduction

In the latter two decades the use of personality measures for personnel selection has undergone a major revival (Viswesvaran, Deller, & Ones, 2007). The development of the Five Factor Model (FFM) (Digman, 1990) has led personality theorists and researchers to re-evaluate the usefulness of personality tests for Industrial and Organizational (I/O) psychology. Meta-analytic procedures have shown that the Five Factors are valid predictors of job performance (Barrick, Mount, & Judge, 2001). Despite the criterion-related validity evidence, the use of personality test scores in selection is not without criticism (Hogan, Barrett, & Hogan, 2007). In applied settings, where subjects are motivated to fake good, response distortion may jeopardize the psychometric properties of the measures (Smith & Robie, 2004). Several studies investigated whether the Five Factor measures retain the same psychometric properties among job applicants. The existing evidence is mixed. Some studies have shown the unbiasedness of the FFM measures in personnel selection and assessment (Montag & Levin, 1994; Tsaousis & Nikolaou, 2001). Other research has provided contrary evidence, suggesting some forms of non-invariance across settings (Livneh & Livneh, 1989; Schmit & Ryan, 1993). Notably, all of these studies have examined configural and metric invariance only. In other words, they have focused on covariance structures, without taking into account the means of the variables.

1.1. Levels of measurement invariance: Assumptions and implications

Configural invariance is the weakest form of invariance. It requires that the examined groups have the same number of factors and the same patterns of fixed and freed factor loadings *per* factor. Metric (or weak) invariance can be assessed by imposing and testing equality constraints on the factor loadings of the observed indicators, across the examined groups. Factor loadings set the metric of the scale by capturing the amount of change that occurs in the observed indicator due to a unit change in the latent factor. When this level of invariance holds, the size of the loadings can be assumed to be approximately the same across groups. This allows one to compare groups in the unstandardized measures of association (e.g. covariances and raw regression weights) between latent factors and external variables.

This level of invariance, however, represents an insufficient condition to meaningfully compare test scores across groups, since the origin of the scale may differ (Byrne & Watkins, 2003). Scalar (or strong) invariance represents a further level of equivalence that can be assessed by imposing equality constraints on the intercepts of the observed indicators. This level of invariance implies that the differences across groups in the means of the observed variables are due to the differences in the means of the respective factors (Steenkamp & Baumgartner, 1998). When scalar invariance is tenable, scores from different groups have the same unit of measurement and the same origin. Accordingly, latent means may be compared across groups (Meredith, 1993). This comports that the observed indicators are unbiased: individuals from different groups with the same score on the underlying trait have the same

* Corresponding author. Tel.: +39 0649917632; fax: +39 0649917711.
E-mail address: michele.vecchione@uniroma1.it (M. Vecchione).

0191-8869/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2011.11.014

expected score on the observed indicator. If scalar invariance does not hold, a statistical comparison of mean scores is meaningless since differences might be due to either real differences or different scale metrics, a condition that in the framework of Item Response Theory is called differential item functioning (Drasgow & Hulin, 1990).

Although invariance at the level of intercepts has relevant implications for meaningful comparisons across groups, we are not aware of any research that used Multiple Group-Confirmatory Factor Analysis (MG-CFA) to assess the equivalence of the intercepts of a FFM measure across applicant and non-applicant samples. As argued by Wu, Li, and Zumbo (2007), this level of invariance has been traditionally neglected by applied psychologists. An extensive review of the measurement invariance literature reveals that 99% of the studies reported tests of invariant factor loadings, whereas only 12% focused on item intercepts (Vandenberg & Lance, 2000).

2. The current study

The current study was designed to assess the measurement invariance of a FFM measure (the Big Five Questionnaire-2, BFQ-2, Caprara, Barbaranelli, Borgogni, & Vecchione, 2007) between applicant and non-applicant groups. This analysis aimed to assess whether the measurement properties of the test are distorted by job applicants' tendencies to convey a positive impression of themselves. Different analytical strategies were employed. First, scale-level analyses were conducted on item composites (i.e. the FFM facets), testing for the equivalence of factor covariance and mean structures. Next, item-level analyses were carried out to further examine the issue of measurement invariance. In this regard, recent studies suggest that the use of item composites as indicators in a CFA model might obscure a lack of measurement invariance for a small number of items (Meade & Kroustalis, 2006). As a consequence, it is recommended to complement the use of scale-level analysis by assessing the equivalence also at the item-level. We conclude by discussing the practical and theoretical implications of results for personality research and assessment.

3. Methods

3.1. Respondents and procedures

Two groups of participants were individually administered the BFQ-2. A first group is composed of 903 respondents (41% males) taken from the general population, who volunteered to participate in the research under anonymous conditions. Participants were recruited as a part of a course assignment in Psychological Statistics at the University of Rome. Mean age of the sample was 35.42 (SD = 9.02); 11% of the sample completed junior high school, and 89% had a university degree. The applicant sample consisted 429 individuals (34% males) applying for business agent jobs for an Italian organization. They took the BFQ-2 as a part of individualized assessment programs, providing their responses during a selection process. Mean age of the sample was 29.88 (SD = 6.95); 14% of the sample completed junior high school, and 86% had a university degree.

3.2. Measures

3.2.1. Big Five Questionnaire-2

The BFQ-2 (Caprara et al., 2007) contains 120 items that form five domain scales (energy, agreeableness, conscientiousness, emotional stability and openness). Respondents indicate agreement with the extent to which each item describes them on a 5-point scale ranging from complete disagreement (very false for me) to complete agreement (very true for me). High correlations between the analogous scales in the BFQ-2 and the NEO-Personality Inventory (Costa & McCrae, 1985) confirm the construct validity of the five domain scales (Caprara et al., 2007).

4. Results

4.1. Preliminary analysis

To reduce the number of indicators, the items of the BFQ-2 were parceled into twenty facets (four for each of the five domains). As recommended (Bandalos & Finney, 2001), we grouped items together that represent similar facets of a construct. Table 1 presents the descriptive statistics and alpha reliability coefficients of both facets and dimensions in the two groups. In accordance with the literature, the observed means on the FFM scales were consistently lower among non-applicants than applicants. These results, according to Cohen's (1988) guidelines, indicate a medium (energy/extraversion, openness) or a large (conscientiousness, agreeableness, emotional stability) difference between the groups. Table 2 presents the intercorrelations among the twenty facets.

4.2. Multi-group confirmatory factor analysis

Since the Five Factor measures are not simple-structured, and have shown inadequate fit when evaluated by confirmatory factor analyses (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996), we employed the "unrestricted factor analysis" (Jöreskog, 1979). This approach takes full advantage of the CFA, allowing examination of different forms of measurement invariance across groups, without assuming that the variables are factorially pure. First, data were examined separately for each group. This is a prerequisite to testing for the equivalence of constructs across groups (Byrne, Shavelson, & Muthén, 1989). Next, metric and scalar invariance were tested across applicant and non-applicant samples using MG-CFA. Results are summarized in Table 3.

4.2.1. Single-group models

As can be observed, the hypothesized Five Factor structure was tenable in both groups (Table 3). The Five Factors were moderately correlated, with an average correlation of 0.27 (SD = 0.13) in the applicant group, and 0.23 (SD = 0.09) among non-applicants.

4.2.2. Model 1: Configural invariance

The posited model fits the empirical data. All the primary loadings were significant, with a mean standardized coefficient of 0.66 within each group. The means of the secondary loadings ranged from 0.13 (non-applicants) to 0.15 (applicants). Subsequent models tested the equivalence of factor loadings and intercepts of the latent factors. The chi-square difference test was used to examine the tenability of the constraints imposed. Since this test suffers from the same limitation as the chi-square test for evaluating overall fit (Steenkamp & Baumgartner, 1998), the invariance hypothesis has been further investigated by considering the differences in the Comparative Fit Index (CFI) between constrained and unconstrained models. In this regard, Cheung and Rensvold (2002) have argued that the invariance hypothesis cannot be retained when the change in CFI is greater than 0.01.

4.2.3. Model 2: Metric invariance

The equality constraints on factor loadings yielded a significant increase of the chi square, $\Delta\chi^2(15) = 32.44$, $p < 0.01$. The significance of results, however, may be due to the large number of subjects considered. The fit indices showed an adequate fit to the