Personality and Individual Differences 55 (2013) 739–743

Contents lists available at SciVerse ScienceDirect

Personality and Individual Differences

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

Factor structure and measurement invariance of a short measure of the Big Five personality traits ☆

Olivier Laverdière [a,*], Alexandre J.S. Morin [b], France St-Hilaire [a]

[a] University of Sherbrooke, Sherbrooke, Canada
[b] Centre for Positive Psychology and Education, University of Western Sydney, Australia

ARTICLE INFO

Article history:
Received 4 January 2013
Received in revised form 6 April 2013
Accepted 3 June 2013
Available online 29 June 2013

Keywords:
Personality traits
Big Five
Validity
Factor structure
Measurement invariance
Psychological assessment
Psychometrics

ABSTRACT

The main purpose of this study is to assess the factor structure and the measurement invariance of the Mini-International Personality Item Pool (Mini-IPIP; Donnellan, Oswald, Baird, & Lucas, 2006). The Mini-IPIP is a brief instrument evaluating personality traits according to the Big Five models. Two samples were collected comprising nearly 800 participants. Confirmatory factor analyzes revealed a five-factor solution consistent with the Big Five model. Measurement invariance analyses showed that the Mini-IPIP was reasonably invariant across samples, genders and age groups. Overall, results pointed to a satisfactory factorial structure and an adequate invariance of the measure.

1. Introduction

The Big Five is the dominant model used to study normal personality across the life span in trait psychology. This model has fuelled a large body of research exploring the validity and relevance of its five-factors: extraversion, agreeableness, conscientiousness, neuroticism, and intellect/openness. It is generally accepted that personality traits are relatively stable, albeit changes are observed over the life course (Marsh, Nagengast, & Morin, in press; Roberts, Walton, & Veichtbauer, 2006). The cross-cultural stability of the FFM has also been the focus of many empirical investigations which generally replicated its factor structure across cultural groups (e.g., McCrae & Allik, 2002).

The Mini International Personality Item Pool (Mini-IPIP) was developed given the widespread interest in the Big Five taking into consideration critical assessment issues, such as questionnaire length (Donnellan et al., 2006). Although the Mini-IPIP has been found to possess promising psychometric properties, such as acceptable reliability and highly similar correlations with other Big Five measures and personality constructs than longer IPIP measures, further investigation of its psychometric properties seems warranted. Importantly, the factorial structure of the Mini-IPIP has not been optimal in previous studies, showing cross-sloadings and elevated correlations between factors that should theoretically be orthogonal (Cooper, Smillie, & Corr, 2010; Donnellan et al., 2006). The issue of factor structure is central to the Big Five approach to personality. While the approach historically rested on factor analysis for the delineation of its main dimensions and the identification of their constituents, poor factorial structure and high correlations among factors have been seen as major shortcomings of the Big Five approach. In this respect, the findings regarding the Mini-IPIP are consistent with previous empirical investigations of this issue with other Big Five measures (e.g., Church & Burke, 1994; Marsh et al., in press; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). Based on confirmatory factor analysis (CFA), these earlier studies usually lead to poor model fit and creative model respecification. However, a main argument leading to the development of the Mini-IPIP was that it would help to overcome these well-documented shortcomings of longer Big-Five personality measures (Donnellan et al., 2006), something that has yet to be empirically demonstrated.

In addition, the measurement invariance of the Mini-IPIP across meaningful subgroups of participants has yet to be investigated. In addition to representing a powerful test of the generalizability of a measurement model across samples and subpopulations, measurement invariance, also represents an important pre-requisite to

☆ This work was supported by research grants from the Canada government (CRSH 767-2004-1917) and Québec government (FQRSC 126223) to Olivier Laverdière.

* Corresponding author. Address: Department of Psychology, University of Sherbrooke, Sherbrooke, Québec J1K2R1, Canada.
E-mail address: olivier.laverdiere@uhsherbrooke.ca (O. Laverdière).

0191-8869/$ - see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.paid.2013.06.008

meaningful and un-biased between-group comparisons. A measure is invariant when it measures the same latent trait across groups in the same manner and with the same precision (e.g. Millsap, 2011). More precisely, the invariance of factor loadings (i.e., weak invariance) tests whether the instrument measures the same construct across subgroups and is a prerequisite to comparisons of latent variances or relations among latent constructs. The invariance of the items' thresholds (i.e. strong invariance) tests whether participants from different subgroups with similar levels on the construct present comparable scores on the items forming the construct and is a prerequisite to latent mean comparisons. Finally, the invariance of the items' uniquenesses (i.e. strict invariance) tests whether the constructs are assessed with similar levels of measurement errors in the various subgroups and is a prerequisite to any group comparison based on manifest (no-latent) scores.

An interesting test of the construct validity of a scale that can easily be combined with tests of measurement invariance had to do with the investigation of potential latent means differences across subgroups in order to verify whether these differences replicate those from previous research investigating the same constructs. For instance, gender and age known to be associated with clear differences in mean-levels of FFM personality traits (Costa, Terracciano, & McCrae, 2001; Donnellan & Lucas, 2008; Feingold, 1994; Lucas & Donnellan, 2009; Roberts et al., 2006; Terracciano, McCrae, Brant, & Costa, 2005). Investigating gender differences in 26 cultures, Costa and colleagues (2001) found that women scored higher on neuroticism, agreeableness, warmth (a facet of extraversion) and openness to feelings, while men were higher on assertiveness (closest to the Mini-IPIP extraversion factor) and openness to ideas (close to the Mini-IPIP intellect/imagination factor). In regard to age differences, results generally show that neuroticism tends to decline with age while conscientiousness increases. Interestingly, these potential latent means differences have yet to be investigated with the Mini-IPIP.

The objectives of the present study are to explore the factorial structure of the Mini-IPIP and to assess its measurement and latent mean invariance according to sample, gender, and age.

2. Method

Two samples were recruited. The first sample was recruited from a large university. Participants were solicited by email for a study on personality (from a pool of approximately 900 individuals who manifested interest to participate in academic research) and 385 gave informed consent. The mean age of this sample was 28.14 years (SD = 9.63), 83% were female, 59% were single, 42% were undergraduate students. The second sample included employees from a large public organization recruited for a study on personality and social relations at work. Approximately 550 employees received an invitation email and 317 gave informed consent. Their mean age was 42.74 years (SD = 10.82), 59% were female, 73% living with a partner. Each of the 20 items from the Mini-IPIP is rated on a five-point Likert scale (ranging from 1 to 5).

Given the ordered categorical nature of Likert scales (Beauducel & Herzberg, 2006; Finney & DiStefano, 2006), all analyses were performed using the robust weighted least square estimator (WLSMV) available in Mplus 7.0 (Muthén & Muthén, 2012). Sample-specific item-level correlations matrices and proportion of respondents using each answer category are available from the first author. As the Big Five model proposes a well-delineated factor structure, CFA models were estimated according to the independent cluster model, with each item allowed to load on a single factor, and all five factors allowed to correlate. The measurement invariance of the final model across subsamples, was tested in the following sequence (Millsap, 2011): (a) configural invariance; (b) loadings (weak) invariance; (c) thresholds (strong) invariance; (d) uniquenesses (strict) invariance; (e) invariance of the a priori correlated uniquenesses; (f) variance–covariance invariance; and (g) latent means invariance. Details on model specification are presented in the appendix (also see Millsap, 2011) and sample inputs are available from the first author. For tests of age-related measurement invariance, age groups were formed based on a median split at age 30, a moment when developmental trends in personality are more constant, compared to early adulthood or retirement age; Marsh et al., 2010; Roberts et al., 2006; Terracciano, Costa, & McCrae, 2006).

It is now broadly accepted that all a priori models will be shown to be false when tested with a sufficiently large sample size. For this and other reasons, chi-square (χ^2) tests of statistical significance are of little relevance for evaluation of goodness of fit and applied CFA research usually predominantly focus on sample size independent indices (Hu & Bentler, 1999; Marsh, Balla, & McDonald, 1988; Marsh, Hau, & Grayson, 2005; Yu, 2002) such as the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), and the root mean square error of approximation (RMSEA). Values greater than 0.90 for CFI and TLI are considered to be indicative of adequate model fit, although values approaching 0.95 are preferable. Values smaller than 0.08 or 0.06 for the RMSEA support respectively acceptable and good model fit. WLSMV χ^2 values are not exact, but rather adjusted to obtain a correct p-value. Thus, WLSMV χ^2 and CFI values can be non-monotonic with model complexity, and χ^2 difference tests need to be conducted via Mplus' DIFFTEST function (MD$\Delta\chi^2$; Asparouhov, & Muthén, 2006). However, these tests tend to be even more problematic than the χ^2 itself as they require additional assumptions (such as the exact fit of the baseline model) that are unlikely to be met (e.g., Marsh et al., 1988). Change in fit indices are thus examined to compare the fit of nested models (Chen, 2007). A ΔCFI of .01 or less and a ΔRMSEA of .015 or less between a more restricted model and the preceding one indicate that the invariance hypothesis should not be rejected. Since indices incorporating a penalty for parsimony (i.e., TLI and RMSEA) can also improve in more restricted models, ΔTLIs were also inspected (Marsh et al., 2005).

3. Results

3.1. Factorial structure

The initial CFA model provides a suboptimal degree of fit to the data (see Table 1, e.g. CFI = .890; TLI = .870; RMSEA = .088). Although the Mini-IPIP does not theoretically possess an intermediate conceptual level between the items and the dimensions, such as the facets seen for longer Big Five instruments, recent findings still suggested that intermediary dimensions may still exist in the IPIP structure (DeYoung, Quilty, & Peterson, 2007). Facing a similar problem in a recent investigation of the NEO-FFI factor structure, Marsh et al. (2010) included correlated uniquenesses between items belonging to unmeasured facets of longer Big Five instruments. This strategy was thus applied for items that had obvious content similarity (#2 and #12, #5 and #20, and #15 and #10). The fit of the model importantly improves up to a satisfactory level with the addition of these three correlated uniquenesses (e.g., CFI = .944; TLI = .932; RMSEA = .064). Standardized loadings from this CFA model are reported in Table 2. None of the standardized loadings were under .300 and only three loadings were under .500 (item 12 from the agreeableness scale: .411; items 5 and 10 from the intellect/imagination scale: respectively .468 and .444) suggesting reasonably well-defined factors for a short measurement scale. Latent factor correlations are reported in Table 3 and show that only one correlation was superior to .30 (.509 between agreeableness and extraversion. The other correlations confirm that the factors are reasonably orthogonal, ranging from −.226 to .273. Table 2 also