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

Article *in* Personality and Individual Differences · October 2013

DOI: 10.1016/j.paid.2013.06.008

3 authors:

Olivier Laverdière
Université de Sherbrooke

Alexandre J S Morin
Concordia University Montreal

France St-Hilaire
Université de Sherbrooke

Volume 55, issue 7, October 2013 ISSN 0191-8869

ELSEVIER

PERSONALITY AND INDIVIDUAL DIFFERENCES

AN INTERNATIONAL JOURNAL OF RESEARCH INTO THE STRUCTURE AND DEVELOPMENT OF PERSONALITY, AND THE CAUSATION OF INDIVIDUAL DIFFERENCES

Contents:

[Continued on outside back cover]

Person. individ. Diff. is indexed/abstracted in:
ASSIA, Curr. Cont. Soc. & Behav. Sci., PASCAL-CNRS Data, Psychol. Abstr, PsycINFO, PsycLIT., Res. Alert, Soc. Sci. Cit. Indx.
Also covered in the abstract and citation database SCOPUS®.
Full text available on ScienceDirect®.

ISSN 0191-8869
55(7) 725–868 (2013)

OFFICIAL JOURNAL OF THE INTERNATIONAL SOCIETY FOR THE STUDY OF INDIVIDUAL DIFFERENCES (ISSID)

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-loadings 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

Table 1
Goodness-of-fit statistics of confirmatory factor analytic (CFA) models.

Model	Description	$\chi^2(df)$	CFI	TLI	RMSEA	90% CI	MD$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA
Main models	1-1. A priori 5-factor model	1029.196 (160)*	.890	.870	.088	.083–.094	–	–	–	–
	1-2. A priori 5-factor model with $r\delta$s	603.471* (157)*	.944	.932	.064	.059–.069	198.221 (3)*	+.054	+.062	−.024
Invariance across samples	2-1. Configural	769.042 (314)*	.945	.933	.065	.059–.070	–	–	–	–
	2-2. Weak (λs)	770.741 (329)*	.945	.938	.062	.056–.068	15.165 (15)	.000	+.005	−.003
	2-3. Strong (λs, vs)	902.713 (384)*	.937	.937	.062	.057-.068	166.358 (55)*	−.008	−.001	.000
	2-4. Strict (λs, vs, δs)	947.677 (404)*	.934	.938	.062	.057-.067	62.675 (20)*	−.003	+.001	.000
	2-5. λ, vs, δs and $r\delta$s	977.149 (407)*	.931	.935	.063	.058–.069	45.313 (3)*	−.003	−.003	+.001
	2-6. λ, vs, δs, $r\delta$s, and ξ/φ	856.341 (422)*	.947	.952	.054	.049–.060	22.801 (15)	+.016	+.017	−.009
	2-7. λ, vs, δs, $r\delta$s, ξ/φ, and ηs	977.977 (427)*	.933	.940	.061	.056-.066	55.923 (5)*	−.014	−.008	+.007
Invariance across age groups	3-1. Configural	808.534 (314)*	.937	.924	.068	.062–.074	–	–	–	–
	3-2. Weak (λs)	808.626 (329)*	.939	.930	.065	.060–.071	14.948 (15)	+.002	+.006	−.003
	3-3. Strong (λs, vs)	910.022 (384)*	.933	.934	.063	.058–.069	125.627 (55)*	−.006	+.004	−.002
	3-4. Strict (λs, vs, δs)	954.238 (404)*	.930	.934	.063	.058–.068	62.789 (20)*	−.003	.000	.000
	3-5. λ, vs, δs and $r\delta$s	957.067 (407)*	.930	.935	.063	.058–.068	5.636 (3)	.000	+.001	.000
	3-6. λ, vs, δs, $r\delta$s, and ξ/φ	877.380 (422)*	.942	.948	.056	.051–.062	35.658 (15)*	+.012	+.013	−.007
	3-7. λ, vs, δs, $r\delta$s, ξ/φ, and ηs	958.400 (427)*	.933	.940	.060	.055-.066	41.331 (5)*	−.009	−.008	+.005
Invariance across genders	4-1. Configural	691.684 (314)*	.952	.941	.059	.053–.065	–	–	–	–
	4-2. Weak (λs)	688.552 (329)*	.954	.947	.056	.050–.062	14.991 (15)	+.002	+.006	−.003
	4-3. Strong (λs, vs)	749.874 (381)*	.953	.953	.053	.047–.058	77.187 (52)*	−.001	+.006	−.003
	4-4. Strict (λs, vs, δs)	778.255 (401)*	.952	.954	.052	.047–.057	39.852 (20)*	−.001	+.001	−.001
	4-5. λ, vs, δs and $r\delta$s	786.284 (404)*	.951	.954	.052	.047–.058	13.440 (3)*	−.001	.000	.000
	4-6. λ, vs, δs, $r\delta$s, and ξ/φ	709.609 (419)*	.963	.966	.045	.039–.050	15.773 (15)	+.012	+.012	−.007
	4-7. λ, vs, δs, $r\delta$s, ξ/φ, and ηs	822.324 (424)*	.949	.954	.052	.047–.057	51.580 (5)*	−.014	−.012	+.007

Note: *$p < .05$; χ^2: WLSMV Chi-square; df: degrees of freedom; CFI: comparative fit index; TLI: Tucker–Lewis index; RMSEA: root mean square error of approximation; 90% CI: 90% confidence interval of the RMSEA; λ: Loadings; v: Thresholds; δ: Uniquenesses; $r\delta s$: a priori Correlated Uniquenesses; ξ: Variance; φ: Covariance; η: Factor means; MD$\Delta\chi^2$: Mplus DIFFTEST χ^2 difference tests; Δ: Change from previous model.

reports scale score reliability coefficients varying from .625 (intellect/imagination) to .835 (conscientiousness). Given the limited number of items used to measure each broad dimension of the Big-Five, and the known association between scale score reliability the number of items in a scale (Sijtsma, 2009), these coefficients are promising. However, the lowest coefficient obtained for intellect/imagination suggests that research based on this instrument would do well to rely on latent variable methodologies allowing for a control of measurement errors (Bollen, 1989).

3.2. Measurement invariance

From this initial model, tests of measurement invariance across samples (Models 2-1 to 2-7), age groups (Model 3-1 to 3-7) and genders (Model 4-1 to 4-7) were then conducted. The results from these tests are reported in Table 1. It should be noted that the answers to three of the intellect/imagination items had to be recoded from their 5 categories answer scales into 4 categories through collapsing the two lowest categories (showing that participants had a very low level of "5-vivid imagination", "15-*difficulty understanding abstract ideas*" and "20-*do not have a good imagination*". Indeed, an important assumption of models based on ordered-categorical items is that the same number of answer categories is used in all groups, an assumption that is violated when there are empty cells due to one specific answer categories not being used in a specific group. Empty cells at the extreme of the answering continuum are common situation in analyses of ordered-categorical items that is classically solved by collapsing of adjacent answer categories (Caci, Morin, & Tran, in press; Reise, Morizot, & Hays, 2007). In order to ensure that no bias results from this procedure, we ensured that all other results were fully replicated with these recoded items. However, only the gender-based invariance tests are based on this new coding scheme.

Examination of the results reported in Table 2 reveals that the complete measurement invariance of the MINI-IPIP factor model is fully supported across sample, age-groups, and genders. Indeed, constraining the factor loadings, items' thresholds, items' uniquenesses, correlated uniquenesses, and even the full latent variance–covariance matrix never resulted in a decrease in fit exceeding the recommended cut-off score of .01 for the CFI and .015 for the RMSEA. In fact, constraining the full variance–covari-

Table 2
Standardized coefficients from the final confirmatory factor analysis model.

Items	λ (unstandardized)	λ (standardized)	h^2
Extraversion			
1. Am the life of the party	1	0.652	0.425
11. Talk to a lot of different people at parties	1.229	0.726	0.527
6. Don't talk a lot	1.069	0.676	0.457
16. Keep in the background	2.615	0.914	0.835
Agreeableness			
2. Sympathize with others' feelings	1	0.559	0.312
12. Feel others' emotions	0.670	0.411	0.169
7. Am not interested in other people's problems	1.645	0.743	0.551
17. Am not really interested in others	3.915	0.935	0.874
Conscientiousness			
3. Get chores done right away	1	0.582	0.336
13. Like order	1.323	0.687	0.473
8. Often forget to put things back in their proper place	1.725	0.777	0.604
18. Make a mess of things	3.341	0.923	0.851
Neuroticism			
4. Have frequent mood swings	1	0.793	0.629
14. Get upset easily	0.986	0.789	0.622
9. Am relaxed most of the time	0.513	0.556	0.309
19. Seldom feel blue	0.585	0.606	0.367
Intellect/imagination			
5. Have a vivid imagination	1	0.468	0.219
10. Am not interested in abstract ideas	0.936	0.444	0.197
15. Have difficulty understanding abstract ideas	1.199	0.536	0.287
20. Do not have a good imagination	1.908	0.711	0.505

All coefficients significant at $p < .05$; λ: Loading; h^2: Communality.

Table 3
Latent correlations and reliabilities from the final confirmatory factor analysis model.

	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Intellect/imagination
Extraversion	0.834	0.509*	0.082	−0.226*	0.273*
Agreeableness		0.770	0.033	−0.171*	0.268*
Conscientiousness			0.835	−.184*	−0.100
Neuroticism				.784	−.157*
Intellect/imagination					.625

Note: $^*p < .05$; scale score reliabilities are reported in the diagonal based on McDonald (1970) omega: $\omega = (\Sigma|\lambda i|)^2 / ([\Sigma|\lambda i|]^2 + \Sigma\delta ii)$ where λi are the factor loadings and δii, the error variances $(1 - h^2)$.

ance matrix to invariance even resulted in a substantial increase in fit that appears to be related to the far greater parsimony of the completely invariant model. However, models 2-7, 3-7 and 4-7 all suggest the non-invariance of the latent factor means. Although all these models do not necessarily result in a decrease in fit exceeding the cut-off score (i.e. Model 3-7 across age-groups), constraining the latent means to be invariant systematically result in the greatest decrease in fit of all models. Given the substantive interest of latent means difference in terms of supporting the construct validity of the Mini-IPIP, we systematically investigate these differences.

Regarding latent means differences across samples, the results revealed that, when the latent means from the student sample were constrained to zero for identification purposes, the latent means were one half of a SD lower in the organizational group for neuroticism (−.502, $p \leqslant .01$) and one third of a SD higher for conscientiousness (.366, $p \leqslant .01$). None of the other latent means differences were statistically significant. Interestingly, two of the main socio-demographic differences observed between both samples are related to gender (with the university sample comprising a higher proportion of females) and age (with the university sample being generally younger), both of which tend to be associated with clear differences in mean-levels of FFM personality traits. Interestingly, age-related latent means differences completely replicate the differences observed across samples, suggesting that these differences might have been mostly related to age differences. More precisely, when the latent means were fixed to zero in the youngest group, the latent means were one half of a SD lower in the oldest group for Neuroticism (−.500, $p \leqslant .01$) and one quarter of a SD higher for conscientiousness (.250, $p \leqslant .01$). Again these results are in line with the results from previous studies of age-related differences in Big Five personality traits, supporting the construct validity of the Mini-IPIP. Likewise, observed mean-level gender-related differences also strongly support the construct validity of the Mini-IPIP (e.g., Costa et al., 2001). More precisely, when the latent means from the male group were constrained to zero, the latent means for the female group were one half of a SD lower for intellect/imagination (−.535, $p \leqslant .01$), and one quarter to one half of a SD higher for Neuroticism (.431, $p \leqslant .01$), Extraversion (.225, $p \leqslant .05$), and Agreeableness (.389, $p \leqslant .01$).

4. Discussion

This study aimed to validate the factor structure of the Mini-IPIP and to verify the measurement invariance of the resulting factor model across samples, genders, and age groups. The results from the initial CFA analysis were suboptimal. However, following Marsh et al.'s (2010) strategy and adding correlated uniquenesses as method factors to account for unmodeled facets scores resulted in a satisfactory fit of the revised model to the data. Also, the factor solution provided unambiguous support to the a priori Big Five factor structure. This result is promising and may reflect a particular strength of the instrument, since previous studies of Big-Five instruments often reported poor model fit (Cooper et al., 2010; Donnellan et al., 2006) or often needed to incorporate secondary ex post facto modifications to obtain suboptimal indices of fit to the data (e.g., Church & Burke, 1994; McCrae et al., 1996). Of course, Mini-IPIP items were selected in order to reduce dimensions intercorrelations and crossloadings, which may explain the sharper factor structure. In addition, the resulting model proved to be quite stable and proved to be reasonably invariant up to the level of the latent variance–covariance matrix across samples, genders, and age groups. Furthermore, the observed latent means differences across genders and age-groups confirmed the results from previous studies regarding age and gender differences in personality (Costa et al., 2001; Donnellan & Lucas, 2008; Feingold, 1994; Lucas & Donnellan, 2009; Roberts et al., 2006; Terracciano et al., 2005), thus contributing to the construct validity of the Mini-IPIP.

Personality measures are generally used with the implicit assumption that scores should not be affected by irrelevant characteristics associated to demographic characteristics or other key variables (e.g., clinical vs. nonclinical populations). However, while drawing conclusions about group differences, the test of this implicit assumption is generally overlooked. The current findings are important in this regard, as the Mini-IPIP was shown to be invariant according to age and gender, two important demographic variables. In this light, the Mini-IPIP appears as a promising instrument for personality traits research.

Overall, the results suggest that the Mini-IPIP possesses a satisfactory factorial structure and is mostly invariant across samples, age and gender. However, one sample included a majority of women (around 80%), a small drawback in terms of generalizability. However, this sample is comparable to the original sample (79%) of Donnellan and colleagues (2006) and this limitation is also offset by evidence of measurement invariance across genders. Future work should look at the cultural invariance of the Mini-IPIP by comparing its structure in culturally diversified samples. An important contribution of the present study is also to present the first evidence regarding the psychometric properties of the French version of the MINI-IPIP. This is particularly interesting for cross cultural studies as French is the official language in 32 countries and territories worldwide (Organisation Internationale de la Francophonie, 2012), including five European countries (France, Belgium, Switzerland, Monaco, and Luxembourg) and Canada, is one of the European institutions' United Nations' official languages and remains the most often taught second language worldwide.

Appendix A

Configural invariance is tested in a model where: (i) items' uniquenesses are fixed to one in the first group and free in the comparison group; (ii) factor means are fixed to zero in the first group and free in the comparison group; (iii) all loadings are freely estimated; (iv) the factor variances are all fixed to 1; (v) the first two thresholds for one referent variable per factor and the first threshold all variables were fixed to equality across groups.

Weak invariance is tested by the addition of equality constraints on the factor loadings across groups, which allows the factors' variances to be freely estimated in the comparison group.

Strong invariance is tested by adding equality constraints on all thresholds across groups. Thresholds reflect the points on the continuous latent response underlying the observed categorical item at which the observed scores change from one category to another.

Strict invariance is tested by adding equality constraints on items' uniquenesses across groups (i.e. fixing them to one in all groups with WLSMV estimation).

When *correlated uniquenesses* are included in a model, their invariance across groups can also be tested through the inclusion of invariance constraints.

Invariance of the factor variance/covariance matrices is tested by adding equality constraints on the factor covariances and by fixing all factor variances to one in all groups and failure of this step does not represent a measurement bias but can be interpreted as a meaningful group difference.

Latent mean invariance is tested by constraining factor means to equality across groups (i.e. fixed to zero in all groups) and failure of this step does not represent a measurement bias but can be interpreted as a meaningful group difference.

References

Asparouhov, T., & Muthén, B. O. (2006). *Robust chi-square difference testing with mean and variance adjusted test statistics*. Los Angeles, CA: Muthén and Muthén. <http://www.statmodel.com/examples/webnote.shtml-web10>.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186–203.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Caci, H. M., Morin, A. J. S., & Tran, A. (in press). Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*. http://dx.doi.org/10.1177/1087054712473834 [in press].

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling, 14*, 464–504.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three and four-dimensional models. *Journal of Personality and Social Psychology, 66*, 93–114.

Cooper, A. J., Smillie, L. D., & Corr, P. J. (2010). A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale. *Personality and Individual Differences, 48*, 688–691.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880–896.

Donnellan, M., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*, 558–566.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18*, 192–203.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429–456.

Finney, S. J., & DiStefano, C. G. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: IAP.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Lucas, R. E., & Donnellan, M. (2009). Age differences in personality: Evidence from a nationally representative Australian sample. *Developmental Psychology, 45*, 1353–1363.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 102*, 391–410.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (in press). Testing measurement invariance and latent mean differences of the Big-Five factor structure over age, gender, and their interaction: Exploratory structural equation modeling tests of plasticity, maturity and La Dolce Vita effects. *Developmental Psychology* [in press].

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275–340). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. S., Trautwein, U., et al. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471–491.

McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York: Kluwer Academic/Plenum Publishers.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating the replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*, 552–566.

McDonald, R. P. (1970). Theoretical foundations of principal factor analysis and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. NY: Taylor & Francis.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.

Roberts, B. W., Walton, K. E., & Veichtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 1–25.

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha [introduction to a special issue]. *Psychometrika, 74*, 107–120.

Terracciano, A., Costa, P. r., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32*, 999–1009.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. r. (2005). Hierarchical linear modeling analyses of the NEO-PI-R Scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20*, 493–506.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.