International Journal of Arts & Sciences,
CD-ROM. ISSN: 1944-6934 :: 09(03):461–472 (2016)

STRUCTURAL VALIDITY OF THE NEO PERSONALITY INVENTORY 3 (NEO-PI-3) IN A FRENCH-CANADIAN SAMPLE

Yann Le Corff and Mathieu Busque-Carrier

Université de Sherbrooke, Canada

The most recent version of the widely used NEO Inventories, the NEO-PI-3, has been translated in several languages worldwide, including Canadian French. Probably the main criticism of the validity NEO-PI-3 has to do with the confirmation of its factor structure in confirmatory factor analysis (CFA). The present study aimed to assess the structural validity of the NEO-PI-3 at the full model level, testing the original model as well as alternative more complex models. Single-factor models at the trait level, and at the basic measurement (facet) level, were also tested. French-Canadian data from real-life settings were analyzed (n = 1313). None of the three full models tested with CFA reached acceptable fit. However, acceptable fit for the oblique full model was reached in the unrestricted framework of partial confirmatory factor analysis (PCFA). Yet, the PCFA revealed substantive crossloadings between Extraversion and Agreeableness facets. Only single-factor models of Neuroticism and Conscientiousness evidenced good model fit, and thirteen of the thirty facets showed acceptable fit. Further analyses indicated that poor model fit for the facets could be explained by high covariances between item measurement errors. The inadequate model fit of the NEO-PI-3 seems to be due to the complexity of the model and to model misspecifications.

Keywords: NEO-PI-3, Five-factor model, Factor Analysis, Structural validity, Measurement model.

Introduction

The NEO Personality Inventory 3 (NEO-PI-3; McCrae & Costa, 2010; McCrae, Costa, & Martin, 2005) is the most recent version of the NEO inventories. It is highly similar to its previous version, the NEO-PI-R, as only 37 out of its 240 items were revised or replaced (McCrae et al., 2005). The NEO-PI-3 has been translated in several languages and is widely used across the world.

The NEO Inventories are designed to measure the Five-Factor Model of personality (FFM; Costa & McCrae, 1992), or Big Five (Goldberg, 1990). These five factors, or broad personality traits, are known as Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. In the NEO-PI-R and 3, each trait is subdivided in 6 facets (for a total of 30 facets). Satisfactory replication of the five-factor structure of the NEO-PI-R/3 in several different languages was achieved using exploratory factor analysis (EFA) with oblique rotation and with the 30 facets as indicators (e.g. Fountoulakis et al., 2014; McCrae, Costa, & Martin, 2005; McCrae, Terracciano, & 78 members of the Personality Profiles of Cultures Project, 2005; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Vassend & Skrondal, 2011), including in French (McCrae, Terracciano et al., 2005; Rolland, Parker, & Stumpf, 1998).

Probably the main criticism of the validity of the NEO Inventories, however, had to do with the confirmation of their factor structure in confirmatory factor analysis (CFA; Gignac, Bates, & Jang, 2007;

Vassend & Skrondal, 2011). Indeed, despite good replication in EFA, several CFA studies failed to support the structural validity of the NEO-PI-R and 3 (e.g. Fountoulakis et al., 2014; Gignac, 2009; Herrmann & Pfister, 2013; Hopwood & Donnellan, 2010; McCrae et al., 1996; Vassend & Skrondal, 2011). The poor model fit in CFA is of particular concern because it raises questions not only in regard to the structural validity of the NEO-PI-3, but also in regard to the validity of its measurement model.

A measurement model represents the relationships between multivariate indicators and their corresponding latent dimensions (Estabrook, 2015). In other words, the measurement model of the NEO-PI-3 provides the rules by which item responses are scored to assess the theoretical constructs. These rules stipulate that each of the 240 items is scored on a single corresponding facet and in turn, each facet is scored on a single corresponding trait. The poor model fit shown in CFA indicate that when indicators (facets) are allowed to load only on their corresponding factor (trait), as stipulated by the measurement model, the model does not fit the data. These results suggest that the measurement model is misspecified (Gignac et al., 2007; Vassend & Skrondal, 2011), and that it does not adequately represent the complexity of the interrelations between personality traits and facets.

Some authors, on the other hand, argue that the complexity of the factor structure and the limitations of CFA make the CFA method inadequate to assess the structural validity of the NEO Inventories (Marsh et al., 2010; McCrae et al., 1996). It is also argued that EFA and exploratory structural equation modeling would be more appropriate because these methods allow indicators to correlate freely with more than one factor (crossloadings).

In this line of thoughts, some researchers (Aluja, Garcia, Garcia, & Seisdedos, 2005; Denis, Crevier-Braud, & Boudrias, 2013; Gorostiaga, Balluerka, Alonso-Arbiol, & Haranburu, 2011) conducted CFA in which some parameters were freed (secondary loadings of indicators were allowed on factors other than the one specified by the measurement model) and obtained statistical models that were judged satisfactory (using fit indexes thresholds less stringent than Hu and Bentler's (1999) "gold standard"). One problem with such a method is that it prevents clear attribution of the indicator to the latent variable it is expected to measure (Fountoulakis et al., 2014), and the resulting model departs from the measurement model. It changes the meaning of observed scores and makes them more complex to interpret (Herrmann & Pfister, 2013). It was also shown that this increased complexity could reduce the convergent and discriminant validity of the model (Herrmann & Pfister, 2013).

It can be maintained that the measurement model of the NEO-PI-3, and even the FFM, may not be an adequate representation of the covariance between the 30 facets, and thus, that the problem is not due to limitations of the CFA method but to the poorly understood complexity of the instrument and to the (mis)specification of the model (Gignac et al., 2007; Vassend & Skrondal, 2011).

A less regarded validity issue is the structural integrity of each individual trait and facet (Gignac et al., 2007). Indeed, only two studies have tested single-factor models of each individual trait (Gignac, 2009; Vassend & Skrondal, 2011). In both studies, CFA results showed good model fit for Neuroticism and Conscientiousness but a lack of model fit for Extraversion and Agreeableness. Openness, however, showed good model fit in Vassend and Skrondal (2011) study and inadequate model fit in Gignac (2009). The inadequate model fit of some of the individual traits makes it improbable for the full model to be found associated with adequate model fit (Gignac, 2009).

Moreover, the validity of the basic measurement model of the NEO, that is, the relationship between item-level responses and the lower-order latent variables (facets) seem to have been taken for granted (Vassend & Skrondal, 2011), even in the NEO manual itself (McCrae & Costa, 2010). To our knowledge, only one study has tested the single-factor models of the facets of the NEO-PI-R (Vassend & Skrondal, 2011). Twenty-three out of the 30 facets reached Hu and Bentler's (1999) suggested cutoffs on at least one relative fit index (CFI or TLI) and one absolute fit index (RMSEA or SRMR). Thus, these results suggest that for the most part, the basic measurement level of the NEO-PI-R appears to be adequate and that the most important misspecification of the model may reside in the relationship between the facets and the traits.

The present study aimed to assess the structural validity of the NEO-PI-3. Analyses were conducted on the five-factor model based on the covariance matrix of the 30 facet scales (the "full model"), as well as on single-factor models at the trait level, and at the basic measurement (facet) level.

Method

Participants

The sample included 420 men and 893 women (n = 1313), aged between 18 and 64 years (M = 33.30, SD = 9.87). It consisted of respondents from the database of the Canadian distributor that completed the Internet version of the French-Canadian NEO-PI-3. The sample thus provided data from real-life assessment settings (mostly personnel selection, organizational psychology, vocational counseling, coaching and psychotherapy assessment).

Instrument

The NEO Personality Inventory 3 (McCrae & Costa, 2010; McCrae, Costa, & Martin, 2005) comprises 240 self-descriptive items with a five-point Likert response format. It was designed to measure the traits of the Five-Factor Model of personality (Neuroticism, Extroversion, Openness, Agreeableness, and Conscientiousness), as well as 30 facets, six per trait, that compose the five higher-order traits. The NEO-PI-3 was translated into Canadian French by the Canadian editor, the Institute of Psychological Research, using a backward translation method with a revision by an expert committee.

Data analysis

Confirmatory factor analyses (CFA) were computed using Mplus (Muthén & Muthén, 1998-2012) and the exploratory factor analysis (EFA) was computed using SPSS. For factor analyses performed on the full model and on single-factor models of individual traits, observed variables were the 30 facets scales (6 per trait). Scores for the 30 facet scales showed sufficient normality according to Muthén and Kaplan's (1985) criterion. For CFA performed on single-factor models of individual facets, observed variables were the items (8 per facet). In models where observed variables were the facet scales, CFA were conducted using the maximum likelihood estimation. In models where observed variables were the items, CFA were carried out using the Robust Mean and Variance Adjusted Weighted Least Squares (WLSMV) estimator method based on polychoric correlations because it is considered the best option for modeling ordinal variables (Brown, 2006), such as the 5-point Likert scales used in NEO-PI-3 items. Model fit was examined using the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square of approximation (SRMR). These indices were preferred because they provide an evaluation of model fit that does not rely strongly on sample size. According to Hu and Bentler's (1999) generally accepted cutoffs, CFI and TLI values of 0.95 or higher, RMSEA values of 0.06 or lower, and SRMR values of 0.08 and lower are considered acceptable. However, Kenny and McCoach (2003) showed that in models with higher number of observed variables CFI and TLI tend to decline and RMSEA tends to improve. According to Hair, Black, Babin, and Anderson (2010), when testing models with high number of observed variables ($m \geq 30$ as is the case when the 30 facets are used as observed variables in the full model), CFI and TLI values of 0.90 or higher, combined with RMSEA values lower than 0.07 and SRMR values of 0.08 and lower, are considered acceptable when sample size is 250 or higher.

Results

Three full models were tested through CFA. Table 1 shows the CFA model fit indexes for each of the three models. The first tested model was the orthogonal simple structure model. According to the measurement model of the NEO-PI-3 and to the theory of the Five-Factor Model of personality (FFM), the five traits are expected to be independent (orthogonal). Also, the measurement model of the NEO-PI-3 implies that each facet scale should be exclusively associated to its corresponding trait scale (Gignac, 2009). Results showed that all indexes used indicated poor model fit.

Previous studies demonstrated that traits of the FFM are in fact not independent but correlated (oblique). Thus, the second model tested was a simple oblique five-factor model. Again, all fit indexes revealed that the model did not provide an acceptable fit to the data.

When the simple CFA models failed to reach acceptable fit, an EFA with oblimin rotation was computed using maximum likelihood estimation and with number of factors to extract fixed at five[1]. This analytic procedure can be qualified as a partial confirmatory factor analysis (PCFA; see Gignac, 2009). Loadings from this PCFA (shown in Table 2) were used to test the third CFA model. This model consisted in an oblique five-factor structure in which parameters were freed according to the factor loadings estimated by PCFA (as suggested in Vassend & Skrondal, 2011). All secondary loadings equal or greater than $\left|.30\right|$ (n = 17) were freed and smaller loadings were set at zero. Again, none of the indexes indicated acceptable model fit.

As shown in Table 1, goodness of fit indexes for the oblimin-rotated maximum likelihood PCFA indicated an acceptable fit according to Hair et al. (2010) less stringent cutoffs for models with higher number of observed variables.

Table 1. Goodness of fit indices for NEO-PI-3 models

Model	x^{2*}	df	CFI	TLI	SRMR	RMSEA
0. Null model	22405.70	435	.00	.00	.29	.20 (.19-.20)
1. Orthogonal five-factor model	9578.07	405	.58	.55	.22	.13 (.13-.13)
Oblique five-factor model						
2. Simple model	7463.78	395	.68	.65	.13	.12 (.11-.12)
3. Modest loadings (≥ .30)	4324.14	378	.82	.79	.10	.09 (.09-.09)
4. Five-factor PCFA oblique model	1801.17	295	.93	.90	.08	.06 (.06-.06)

Note. x^2 = chi-square; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation and its 90% confidence interval.
* The associated *p* values were always lower than .001

Table 3 shows goodness of fit indexes for each of the five NEO-PI-3 latent variables at the trait level. While SRMR indicated acceptable model fit for each of the five single-factor models, RMSEA indicated unacceptable model fit for each model. This discrepancy between the two absolute fit indexes may be due to the fact that RMSEA is biased upward in models with low degrees of freedom (Kenny, Kaniskan, & McCoach, 2014). Both CFI and TLI indicated unacceptable fit for Extraversion, Openness and Agreeableness. However, CFI coefficients for Neuroticism and Conscientiousness indicated good model fit and TLI coefficients were very close to reaching the threshold for good model fit. It is worth noting

[1] Preliminary to this analysis, a principal component analysis with varimax rotation was conducted. Both Horn's parallel analysis (95th percentile) and Catell's scree test strongly suggested a five factor structure.

that Marsh, Hau and Wen (2004) argue that a TLI cutoff of 0.95 may be too stringent and leads to higher percentage of incorrect rejection of specified models (larger type 1 errors) than other incremental fit indexes such as the CFI. Thus, it can be considered that the present results support the model fit for Neuroticism and Conscientiousness.

Table 2. Oblimin-rotated maximum-likelihood factor loadings (pattern matrix) for the NEO-PI-3 model

	N	E	O	A	C
Neuroticism facets					
N1 : Anxiety	**-.84**	-.03	.01	.12	.03
N2 : Angry Hostility	**-.71**	-.10	.00	**-.34**	-.09
N3 : Depression	**-.78**	-.16	.06	.09	-.16
N4 : Self-Consciousness	**-.63**	-.27	-.01	.20	-.12
N5 : Impulsiveness	**-.57**	.14	.00	**-.31**	**-.31**
N6 : Vulnerability	**-.66**	-.11	-.07	.13	**-.30**
Extraversion facets					
E1 : Warmth	.07	**.80**	.09	.15	.08
E2 : Gregariousness	.11	**.68**	-.12	-.06	-.01
E3 : Assertiveness	**.31**	**.37**	.01	**-.49**	.26
E4 : Activity	.06	**.34**	-.05	**-.42**	**.33**
E5 : Excitement-Seeking	-.01	**.30**	.12	**-.40**	-.02
E6 : Positive Emotions	.09	**.63**	.12	-.03	.03
Openness facets					
O1 : Fantasy	-.18	.11	**.54**	-.12	-.25
O2 : Aesthetics	-.08	-.06	**.68**	.04	.02
O3 : Feelings	**-.44**	**.37**	**.44**	-.10	.08
O4 : Actions	**.32**	.15	**.32**	-.11	-.16
O5 : Ideas	.18	-.18	**.68**	-.09	.22
O6 : Values	.03	.07	**.43**	.05	-.14
Agreeableness facets					
A1 : Trust	**.36**	**.39**	.03	**.18**	.00
A2 : Straightforwardness	.03	.16	-.06	**.54**	.19
A3 : Altruism	.01	**.56**	.11	**.37**	.27
A4 : Compliance	.27	.06	-.02	**.64**	-.01
A5 : Modesty	-.20	.01	-.07	**.51**	-.08
A6 : Tender-Mindedness	-.13	.28	.29	**.36**	.12
Conscientiousness facets					
C1 : Competence	**.40**	.08	.13	-.02	**.57**
C2 : Order	-.08	-.03	-.10	-.04	**.66**

C3 : Dutifulness	.01	.10	.04	.18	**.72**
C4 : Achievement Striving	.01	.15	.02	-.20	**.72**
C5 : Self-discipline	.26	.13	-.07	-.03	**.68**
C6 : Deliberation	.19	-.22	.07	**.32**	**.58**
% of variance	27.90 %	10.39 %	3.76 %	7.49 %	4.81 %

Note. Factor loadings of 0.30 or more are given in boldface.

Table 3. Goodness of fit indices for the five NEO-PI-3 traits

One-factor model	x^{2*} (df = 9)	CFI	TLI	SRMR	RMSEA (90% CI)
Neuroticism	222.76	.96	.93	.04	.13 (.12-.15)
Extraversion	370.15	.86	.76	.06	.18 (.16-.19)
Openness	207.56	.85	.75	.06	.13 (.12-.15)
Agreeableness	249.25	.80	.67	.06	.14 (.13-.16)
Conscientiousness	192.73	.95	.92	.04	.13 (.11-.14)

Note. x^2 = chi-square; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation and its 90% confidence interval.
* The associated *p* values were always lower than .001

A closer inspection of the factor loadings from the oblimin-rotated maximum likelihood PCFA helps explain CFA results. As shown in Table 2, all facets from Neuroticism and Conscientiousness traits have a strong primary loading (at least ±0.57) on their corresponding factor. Neuroticism and Conscientiousness are the two traits that reached acceptable model fit in single-factor CFA. Openness facets all had their primary loading on the Openness factor, but three out of six facets had a loading lower than 0.45, which can be qualified as less than "fair" (Tabachnick & Fidell, 2013). Extraversion and Agreeableness were more problematic with some of their facets having their primary loading on the other trait. E1 Warmth, E2 Gregariousness, E6 Positive emotions, A1 Trust, and A3 Altruism had their primary loading on the same latent variable. E3 Assertiveness, E4 Activity, E5 Excitement-Seeking, A2 Straightforwardness, A4 Compliance, A5 Modesty, and A6 Tender-Mindedness had their primary loading on another latent variable.

Table 4 shows CFA results at the basic measurement level, that is, the relationship between items and the latent variables at the facet level. Only thirteen of the thirty facets showed good model fit on at least one relative fit index (CFI or TLI) and one absolute fit index (RMSEA or SRMR): N1 Anxiety, N3 Depression, N4 Self-Consciousness, N6 Vulnerability, E1 Warmth, E2 Gregariousness, A2 Straightforwardness, A4 Compliance, A6 Tender-Mindedness, C1 Competence, C3 Dutifulness, C5 Self-discipline, and C6 Deliberation. It is worth noting that Openness facets evidenced particularly poor model fit compared to the other groups of facets.

One reason for poor fit for several of the single-factor models of facets could be the presence of high measurement error covariances between indicators (Byrne, 2010; Gignac et al., 2007). Measurement error covariances represent systematic (rather than random) measurement error in items and are indicated by high modification indices (MI) in CFA. They may derive from items characteristics such as overlap in item content or the presence of a small subfactor within the factor (Byrne, 2010). For example, in the present study, the highest MI was associated with items 3 and 5 of the Ideas facet (MI = 1260.93). In this case, the measurement error covariance can be attributed to the high degree item content overlap, as both items have a very similar wording ("I enjoy […] puzzles").

Table 4. Goodness of fit indices for the 30 NEO-PI-3 facets

	x^{2*} (df = 20)	CFI	TLI	SRMR	RMSEA (90% CI)
Neuroticism facets					
N1 : Anxiety	197.81	.99	.98	.03	.08 (.07-.09)
N2 : Angry Hostility	260.41	.94	.91	.05	.10 (.09-.11)
N3 : Depression	478.76	.96	.94	.05	.13 (.12-.14)
N4 : Self-Consciousness	117.36	.98	.98	.03	.06 (.05-.07)
N5 : Impulsiveness	758.19	.85	.79	.09	.17 (.16-.18)
N6 : Vulnerability	483.34	.95	.93	.06	.13 (.12-.14)
Extraversion facets					
E1 : Warmth	277.12	.95	.93	.05	.10 (.09-.11)
E2 : Gregariousness	293.73	.95	.93	.05	.10 (.09-.11)
E3 : Assertiveness	372.44	.93	.90	.06	.12 (.11-.13)
E4 : Activity	716.60	.83	.76	.09	.16 (.15-.17)
E5 : Excitement-Seeking	236.87	.91	.88	.06	.09 (.08-.10)
E6 : Positive Emotions	381.36	.92	.88	.07	.12 (.11-.13)
Openness facets					
O1 : Fantasy	798.20	.79	.71	.10	.17 (.16-.18)
O2 : Aesthetics	822.09	.89	.84	.08	.18 (.17-.19)
O3 : Feelings	237.87	.90	.86	.06	.09 (.08-.10)
O4 : Actions	425.00	.81	.73	.08	.12 (.11-.14)
O5 : Ideas	1706.77	.76	.66	.13	.25 (.24-.26)
O6 : Values	166.49	.90	.85	.05	.08 (.06-.09)
Agreeableness facets					
A1 : Trust	259.20	.94	.92	.05	.10 (.09-.11)
A2 : Straightforwardness	154.97	.99	.98	.05	.07 (.06-.08)
A3 : Altruism	265.49	.92	.89	.06	.10 (.09-.11)
A4 : Compliance	86.76	.96	.94	.04	.05 (.04-.06)
A5 : Modesty	653.35	.80	.72	.09	.16 (.15-.17)
A6 : Tender-Mindedness	179.56	.95	.94	.05	.08 (.07-.09)
Conscientiousness facets					
C1 : Competence	211.38	.96	.94	.05	.09 (.08-.10)
C2 : Order	340.03	.94	.92	.07	.11 (.10-.12)
C3 : Dutifulness	121.70	.97	.95	.04	.06 (.05-.07)
C4 : Achievement Striving	375.00	.92	.88	.06	.12 (.11-.13)
C5 : Self-discipline	167.78	.99	.98	.03	.08 (.07-.09)
C6 : Deliberation	261.14	.96	.94	.05	.10 (.09-.11)

Note. x^2 = chi-square; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation and its 90% confidence interval.
* The associated *p* values were always lower than .001

MI for all pairs of items of each facet were examined in order to identify high measurement error covariances that may explain poor model fit. For each facet that did not reach good model fit in the above analyses (see Table 4), respecified single-factor models were successively tested in CFA by freeing the path between the pair of error terms with the highest MI until an adequate fit for the model was reached. In addition to having a substantial MI, a measurement error covariance had to be substantively meaningful (e.g. obvious item content overlap) to be added to the model (Byrne, 2010). Following these respecifications, all facets reached acceptable model fit on at least one relative fit index and one absolute fit index. Table 5 shows the CFA results for the 30 final models (17 were respecified) with the number of respecifications (freed paths) needed to reach acceptable model fit.

Table 5. Goodness of fit indices for the 30 NEO-PI-3 facets after model respecifications

	Df	Respec.	X^{2*}	CFI	TLI	SRMR	RMSEA (90% CI)
Neuroticism facets							
N1 : Anxiety	20	0	197.81	.99	.98	.03	.08 (.07-.09)
N2 : Angry Hostility	19	1	183.96	.96	.94	.05	.08 (.07-.09)
N3 : Depression	20	0	478.76	.96	.94	.05	.13 (.12-.14)
N4 : Self-Consciousness	20	0	117.36	.98	.98	.03	.06 (.05-.07)
N5 : Impulsiveness	18	2	134.00	.98	.96	.04	.07 (.06-.08)
N6 : Vulnerability	20	0	483.34	.95	.93	.06	.13 (.12-.14)
Extraversion facets							
E1 : Warmth	20	0	277.12	.95	.93	.05	.10 (.09-.11)
E2 : Gregariousness	20	0	293.73	.95	.93	.05	.10 (.09-.11)
E3 : Assertiveness	19	1	129.93	.98	.97	.04	.07 (.06-.08)
E4 : Activity	18	2	211.45	.95	.93	.05	.09 (.08-.10)
E5 : Excitement-Seeking	18	2	139.28	.95	.92	.05	.07 (.06-.08)
E6 : Positive Emotions	19	1	13.07	.97	.96	.04	.07 (.06-.08)
Openness facets							
O1 : Fantasy	17	3	172.93	.96	.93	.05	.08 (.07-.10)
O2 : Aesthetics	18	2	265.99	.97	.95	.05	.10 (.09-.11)
O3 : Feelings	18	2	132.23	.95	.92	.04	.07 (.06-.08)
O4 : Actions	16	4	104.43	.96	.93	.04	.07 (.05-.08)
O5 : Ideas	18	2	245.38	.97	.95	.05	.10 (.09-.11)
O6 : Values	18	2	85.55	.95	.93	.04	.05 (.04-.07)
Agreeableness facets							
A1 : Trust	19	1	161.34	.96	.95	.04	.08 (.07-.09)
A2 : Straightforwardness	20	0	154.97	.99	.98	.05	.07 (.06-.08)
A3 : Altruism	19	1	16.10	.95	.93	.05	.08 (.07-.09)
A4 : Compliance	20	0	86.76	.96	.94	.04	.05 (.04-.06)
A5 : Modesty	14	6	109.57	.97	.94	.04	.07 (.06-.09)
A6 : Tender-Mindedness	20	0	179.56	.95	.94	.05	.08 (.07-.09)

Conscientiousness facets							
C1 : Competence	20	0	211.38	.96	.94	.05	.09 (.08-.10)
C2 : Order	18	2	211.86	.96	.94	.05	.09 (.08-.10)
C3 : Dutifulness	20	0	121.70	.97	.95	.04	.06 (.05-.07)
C4 : Achievement Striving	19	1	114.12	.98	.97	.04	.06 (.05-.07)
C5 : Self-discipline	20	0	167.78	.99	.98	.03	.08 (.07-.09
C6 : Deliberation	20	0	261.14	.96	.94	.05	.10 (.09-.11)

Note. Respec. = number of respecifications made to the initial model; df = degrees of freedom; x^2 = chi-square; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation and its 90% confidence interval.
* The associated *p* values were always lower than .001

Discussion

The present study assessed the structural validity of the NEO-PI-3 at the full model level, testing the original model as well as alternative more complex models, in a French-Canadian sample. Furthermore, it is the first study to examine single-factor models at the trait level, and at the facet level, in the most recent version of the NEO Inventories.

The results showed that in a PCFA with oblique rotation and with all parameters freed, the five-factor structure had a good fit to the data. In the more restrictive framework of the CFA, however, the model did not show adequate fit, even with an oblique rotation and with all secondary loadings greater than 0.30 freed. These results are consistent with previous reports from other versions of the NEO. As Vassend and Skrondal (2011) remarked, "To obtain at least an approximate model fit, however, practically all of the factor loadings in the factor pattern matrix have to be treated as significant parameters" (p. 1303). Thus, the factor structure obtained from the correlation matrix of the 30 facets appears more complex than what is suggested in the measurement model of the NEO-PI-3 or, in other words, the measurement model does not account for the complexity observed in the data.

It was argued that this lack of model fit could be due to the limitations of the CFA method and to the complexity of the model (Marsh et al., 2010; McCrae et al., 1996). In the light of the present study, the inadequate model fit seems to be due, at least in part, to the misspecifications of the model itself, both at the trait level and at the facet level. Indeed, single-factor models of Extraversion, Openness, and Agreeableness did not show adequate fit to the data. These results suggest that adopting a more complex measurement model for the NEO-PI-3 based on EFA results (e.g. Herrmann & Pfister, 2013) may only hide the problem of misspecification at the trait level. Additionally, at the basic measurement level, approximately half of the facets evidenced inadequate model fit. Further analyses indicated that poor model fit at the basic measurement level could be explained by high covariances between item measurement errors. Adequate model fit was reached for all 30 facets when the substantively meaningful measurement error covariances with the largest MI were added to the models. These results suggest that the rewording of specific items could help reduce item measurement error covariances and thus increase the validity of the model at the basic measurement level.

It is also of interest that EFA results showed that Extraversion and Agreeableness facets did not load according to the theoretical structure of the NEO-PI-3, and that the exact same loading pattern was observed in the Swedish normative sample for the NEO-PI-3 (Källmen, Wennberg, & Bergman, 2011). These results, however, make theoretical sense. Indeed, in the present study, Warmth, Gregariousness, Positive emotions, Trust, and Altruism had their primary loading on the same latent variable which can be interpreted as an Affiliation (or Love-Hostility) trait (see DeYoung, Weisberg, Quilty, & Peterson, 2013; McCrae & Costa, 1989). On the other hand, Compliance, Modesty, Tender-Mindedness, Assertiveness (inverted), Activity (inverted), Excitement-Seeking (inverted), and Straightforwardness (inverted) had their primary loading on the same latent variable which can be interpreted as a Dominance-Submission

trait (see DeYoung Quilty, & Peterson, 2007; McCrae & Costa, 1989; Pincus, 2002). These two latent variables (Affiliation and Dominance) correspond to the two axis of the Interpersonal Circumplex Model of personality (Wiggins, 1979). Extraversion and Agreeableness (of the FFM) have been conceptualised as rotational variants of these two axis (McCrae & Costa, 1989; Pincus, 2002). These results suggest that, in the present French-Canadian sample, Affiliation and Dominance of the Interpersonal Circumplex Model of personality may better represent the correlation patterns of the Extroversion and Agreeableness facets of the NEO-PI-3.

One limitation of this study is that we used strict cutoff scores for the goodness of fit indexes, based on Hu and Bentler (1999) and Hair et al. (2010) recommendations. Marsh et al. (2004) argue that Hu and Bentler's (1999) thresholds for fit indexes are too stringent for complex models in applied research, notably because they based their conclusions on models comprising three five-item factors. According to some authors, in applied research, CFI and TLI between 0.80 and 0.90 could indicate acceptable model fit and indexes over 0.90 would indicate good model fit (e.g. Ivanova et al., 2015). Even with these lowered thresholds, however, the TLI index would still indicate poor model fit for the three full models tested in CFA. Another limitation is that we used the French-Canadian translation of the NEO-PI-3, for which no validity data has yet been published, aside from the present study. It is possible that this translated version is not entirely equivalent to the original English version.

In conclusion, in the present study, confirmatory factor analyses did not support the structural validity of the NEO-PI-3. These results do not necessarily imply that the NEO-PI-3 is not a valid measure of personality. It is likely that the high crossloadings are actual reflections of the reality of the personality structure, where traits and facets are distributed throughout the factor space. It appears, though, that the measurement model of the NEO-PI-3 fails to account for that complexity. Moreover, the poor model fit even at the basic measurement level, and the crossloading pattern between Extroversion and Agreeableness facets, suggest model misspecifications that deserve further attention.

References

1. Aluja, A., Garcia, O., Garcia, L. F., & Seisdedos, N. (2005). Invariance of the "NEO-PI-R" factor structure across exploratory and confirmatory factor analyses. *Personality and Individual Differences, 38,* 1879-1889. doi: 10.1016/j.paid.2004.11.014
2. Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: Guildford.
3. Byrne, B. M. (2010). *Structural Equation Modeling with AMOS* (2nd Edition). New York: Routledge.
4. Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*, 653–665.
5. Denis, P. L., Crevier-Braud, L., & Boudrias, J.-S. (2013). NEO-PI-R: Comparaison de la structure factorielle des versions québécoises et française. [Factor structure comparison of Quebec and French versions of the NEO PI-R] *Revue Québécoise de Psychologie, 34,* 211-241.
6. Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences, 28,* 313-350. doi: 10.1017/S0140525X05000063
7. DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: Ten aspects of the Big Five. *Journal of Personality and Social Psychology, 93,* 880-896. doi: 10.1037/0022-3514.93.5.880
8. DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the Interpersonal Circumplex, and trait affiliation. *Journal of Personality, 81,* 465-475. doi: 10.1111/ jopy.12020
9. Estabrook, R. (2015). Evaluating measurement of dynamic constructs: Defining a measurement model of derivatives. *Psychological Methods, 20,* 117-141. doi: 10.1037/a0034523
10. Fountoulakis, K. N., Siamouli, M., Moysidou, S., Pantoula, E., Moutou, K., Panagiotidis, P., et al. (2014). Standardization of the NEO-PI-3 in the Greek general population. *Annals of General Psychiatry, 13.* Retrieved from http://www.annals-general-psychiatry.com/content/pdf/s12991-014-0036-9.pdf. doi:10.1186/s12991-014-0036-9

11. Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO-PI-R. *Journal of Personality Assessment, 91,* 40-47. doi: 10.1080/00223890802484126
12. Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit, reliability, and the five-factor model as measured by the NEO-FFI. *Personality and Individual Differences, 43*, 1051-1062. doi: 10.1016/j.paid.2007.02.024
13. Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,*1216-1229.
14. Gorostiaga, A., Balluerka, N., Alonso-Arbiol, I., & Haranburu, M. (2011). Validation of the Basque Revised NEO Personality Inventory (NEO PI-R). *European Journal of Psychological Assessment, 27,* 193-205. doi: 10.1027/1015-5759/a000067
15. Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
16. Herrmann, A. & Pfister, H.-R. (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in the Big Five inventories? *Journal of Research in Personality, 47*, 599-608. doi: 10.1016/j.jrp.2013.05.004
17. Hopwood, C. J. & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review, 14*, 332-346. doi: 10.1177/1088868310361240
18. 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. doi: 10.1080/10705519909540118
19. Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Turner, L. V., Ahmeti-Pronaj, A., Au, A., et al. (2015). Syndromes of self-reported psychopathology for ages 18-59 in 29 societies. *Journal of Psychopathology and Behavioral Assessment, 37,* 171-183. doi: 10.1007/s10862-014-9448-8
20. Källmen, K., Wennberg, P., & Bergman, H. (2011). Psychometric properties and norm data of the Swedish version of the NEO-PI-R. *Nordic Journal of Psychiatry, 65,* 311-314. doi: 10.3109/08039448.2010.545433
21. Kenny, D. A., Kaniskan, & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Resarch*. Advance online publication. doi: 10.1177/0049124114543236.
22. Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10,* 333-351. doi: 10.1207/S15328007SEM1003_1
23. Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320-341. doi: 10.1207/s15328007sem1103_2
24. Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471-491. doi: 10.1037/a0019227
25. McCrae, R. R. & Costa, P. T., Jr. (2010). *The NEO Inventories. Professional Manual*. Lutz, FL: Psychological Assessment Ressources.
26. McCrae, R. R., & Costa, P. T., Jr. (1989). The structure of interpersonal traits: Wiggin's circumplex and the Five-Factor Model. *Journal of Personality and Social Psychology, 56,*586-595.
27. McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment, 84,* 261-270. doi: 10.1207/s15327752jpa8403_05
28. McCrae, R. R., Terracciano, A., & 78 members of the Personality Profiles of Cultures Project (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology, 88,* 547-561. doi: 10.1037/0022-3514.88.3.547
29. McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Personality Processes and Individual Differences, 70*, 552-566.
30. Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 431-462.
31. Muthén, L. K., & Muthén, B. (2012). 1998-2012. *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthen & Muthen.

32. Pincus, A. L. (2002). Constellation of dependency within the Five-Factor Model of personality. In P. T. Costa, Jr., & T. A Widiger (Eds.), *Personality Disorders and the Five-Factor Model of Personality* (2nd edition, pp. 203-214). Washington, DC: American Psychological Association.
33. Rolland, J.-P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translations of the NEO-PI-R and NEO-FFI. *Journal of Personality Assessment, 71,* 269-291.
34. Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics (6th Edition)*. Boston: Pearson.
35. Vassend, O., & Skrondal, A. (2011). The NEO personality inventory revised (NEO-PI-R): Exploring the measurement structure and variants of the five-factor model. *Personality and Individual Differences, 50,* 1300-1304. doi: 10.1016/j.paid.2011.03.002
36. Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37,*395-412. doi: 10.1037/0022-3514.37.3.395