Validación de la versión española del NEO-FFI-30

Validation of the NEO-FFI-30 Spanish version

Ascensión Fumero
Facultad de Psicología, Universidad de La Laguna
Instituto Universitario de Neurociencia (IUNE), Universidad de La Laguna

Adelia de Miguel
Facultad de Psicología, Universidad de La Laguna

Resumen
En este estudio se analizaron las propiedades psicométricas de la versión española del Inventario de Cinco Factores NEO de treinta items (NEO-FFI-30) en un total de 2096 adultos (61.55% mujeres) con edades comprendidas entre los 18 y los 71 años. Se analizó la validación cruzada de la estructura factorial mediante un análisis factorial exploratorio de Ejes Principales (PAF) y un Análisis Factorial Confirmatorio (AFC). El PAF reprodujo la estructura original del NEO-FFI-30 y el AFC mostró un ajuste satisfactorio de los modelos monofactoriales para las cinco dimensiones. Los resultados apoyan la hipótesis de que el conjunto abreviado de 30 ítems del FFI puede utilizarse eficazmente para evaluar el modelo de personalidad de cinco factores de forma fiable y válida, y con capacidad convergente, diferencial y concurrente con respecto al malestar clínico y psicológico. La versión española del NEO-FFI-30-SF es un instrumento robusto para medir las cinco dimensiones de la personalidad.

Palabras clave
Modelo de cinco factores de personalidad; versión corta; NEO-FFI-30-SF; malestar psicológico; población española.

Abstract
In this study, the psychometric properties of the Spanish form of the NEO Five-Factor Inventory (NEO-FFI-30) were analyzed on a total of 2096 adults (61.55% female) with age ranged from 18 to 71 years. The cross-validation of the factorial structure was analyzed by an exploratory Principal Axis Factoring (PAF) and a Confirmatory Factor Analysis (CFA). PAF reproduced the original structure of NEO-FFI-30 and CFA showed a satisfactory fit of single-factor models for the five dimensions. The results provided sufficient support for the hypothesis that an abbreviated set of 30 FFI items can be used efficiently to assess the five-factor personality model in a reliably and valid manner. The analyses showed a structure congruent with previous international adaptations of the NEO-FFI-30, with high values of internal consistency, and convergent, differential and concurrent ability with respect to clinical and psychological distress. The NEO-FFI-30-SF is a robust instrument to measure the five dimensions of personality.

Keywords
Five-factor model of personality; short version; NEO-FFI-30-SF; psychological distress; Spanish population.
Introduction

The Big Five Factor Model (McCrae & Costa, 1987) proposed an assessment method of the main personality traits -neuroticism, extraversion, openness to experience, agreeableness and conscientiousness. NEO-PI-R (Costa & McCrae, 1992), which includes the five basic tendencies and 30 facets (6 per tendency), has been generally well accepted by the international psychology community, and applied in many cultures (McCrae, 2017). The Spanish version of this questionnaire (Cordero, Pamos, & Seisdedos, 1999) has been extensively validated (Aluja et al., 2005; Sanz & García-Vera, 2009).

The successive replications of the structure of the 240 items original version in various linguistic and socio-cultural contexts have evidenced however the existence of problems as numerous items have higher loadings on other factors than on the original factor; some items have loadings on different factors that were not observed on their assigned or the orthogonality was not observed (Vassend & Skrondal, 2011). Short versions have been proposed and revised trying to improve the factorial structure, the understanding of the items themselves and avoiding possible biases such as acquiescence (Hahn et al., 2012; Rammstedt & John, 2007).

Currently, many research studies require measures in settings in which assessment-related resources (e.g., time) are limited or information about participants’ personality should be based exclusively on the five factors, but not on the facets. So, it is worth investigating whether a reduced set of NEO items can measure the five factors in an efficient, reliable and valid manner. Previous studies in Spain provide support for the possibility of identifying a smaller set of items from the original version and adequate validated measures that do not result in a significant reduction of reliability or validity. An example of this can be using Spanish language (Aluja et al., 2005) or Basque language (Haranburu et al., 2007), both of them consisting of 60 items.

A briefer version is the NEO-FFI-30 (Körner et al., 2007), which could be an alternative to facilitate a short assessment based on the Big Five model. This version, obtained from NEO-FFI (Costa & McCrae, 1989), replicated the factor structure, and the new short scales showed good internal consistency and were highly correlated with the original NEO−FFI scales. This finding supported the factor and construct validity of this competitive, economic instrument without any significant loss of information.

Since there are no results on the Spanish adaptation of the NEO-FFI-30, the present study aims to provide data to clarify its validation.

In order to create the Spanish Form (NEO-FFI-30-SF), the correlated 30 items from Körner et al. (2007) were extracted from the Spanish version of NEO-PI-R (Cordero et al., 1999) and completed by participants from the general population. The main objective of this study was to confirm the invariance of this Spanish Form of the 30-item NEO-FFI that will allow operationalizing the Big Five model in a shortened way and can be used as a broad-spectrum diagnostic tool that should solve the founded problems in the original factorial structure.
The Big-Five Factor Model and NEO-PI-R have been used in several areas in psychology, one of these areas being mental and physical health. Several examples are planning treatment for personality disorders (Sanderson & Clarkin, 2002), examining the association between personality traits and subjective ratings of mental and physical health (Colodro et al., 2018; Löckenhoff et al., 2008), and analyzing role of personality traits and HIV progression (Ironson et al., 2008). A secondary objective of this study was also intended to test whether the Spanish Form of the short version maintained the convergent, discriminant and predictive ability, like the original NEO-PI-R version, with respect to clinical and psychological distress.

**Methods**

**Participants**

There were 2096 participants (61.55% female) from the general population. Age range was 18 to 71 years ($M_{age} = 24.37$, $SD = 8.22$). There were sex differences ($M_{female} = 23.81$, $SD_{female} = 7.85; M_{male} = 25.38$, $SD_{male} = 8.77; t = -4.01$, $p = .001$, $d = .19$). The employment status was evaluated, with 77.6% college students, 12.12% workers, 0.3% homemakers and 2.77% unemployed. There were no sex differences in this variable ($X^2 = 36.88$, $p < .001$).

This study was carried in two phases. In the first phase, a total of 1622 students from University of La Laguna fulfilled the questionnaire in order to examine the validity of NEO-30-SP items. There were 68.6% males and 31.4% females at the age of 23.94 ($SD = 7.97$). A total of 474 adults participated in the second phase study, 50.2% were male and 49.8% were female at the age of 25.80 ($SD = 8.86$).

**Instruments**

The NEO-PI-R (Costa & McCrae, 1992) is the original 240-item questionnaire that measures the five basic tendencies and their 30 facets of temperament, using a five-point Likert scale (strongly disagree = 0, strongly agree = 4). Spanish version’s alphas ranged from .86 for A to .90 for N (Sanz & García-Vera, 2009). Scores for the five traits to both original (48 items for each one) and short (6 items for each one) versions were calculated.

The SCL-90-R (Derogatis, 1994, Spanish version from González de Rivera et al., 2002) consists of ratings of 90 symptoms on a five-point scale (not at all = 0, extremely = 4) indicating how frequently the participant has experienced these symptoms in the last week. We assessed the nine clinical subscales. Reliabilities for Spanish version ranged from .88 to .81; in the present sample, they ranged from .75 to .89.

The IPDE (Loranger, 1977, Spanish version from de Miguel & Pelechano, 2000) is a screening questionnaire of personality disorders for the International Classification of Disorders (ICD-10, WHO, 1993). It consists of 59 items, which are self-reported by the patient as true = 1 or false = 0, regarding the last five years. The median kappa values for interrater agreement was 0.77 (Loranger, 1977).

**Design**

This ex post-facto study utilized a transversal, correlational and quantitative design.
The incidental snowball sampling technique was used. Participation was voluntary. Participants were informed that they could drop out at any time. There is no data on excluded participants because only surveys with responses to all items were collected.

**Procedure**

In the first phase, participants were recruited among the body of psychology students registered in several courses in personality psychology. In the second phase, the students of a course in personality psychology were asked to answer and to distribute the file via e-mail to relatives and friends in their inner circle. To increase the sample variability, each student was asked to invite 10 of their acquaintances (age range 18–75) to participate. They received a course credit in return for their collaboration. Confidentiality and anonymity were ensured in all participants.

**Statistical Analysis**

The statistical procedure included: in the first phase, (1) an exploratory factor analysis of NEO-FFI-30 tested via SPSS 22.0; in the second phase (2) an confirmatory factor analysis of NEO-FFI-30 tested via AMOS 5.0; (3) trait score differences between original and short versions by test-retest analysis; (4) gender differences in both original and short versions by t and size effect (Cohen’s d); and (5) Pearson’s r among the five personality factors (both original and short versions) and clinical symptomatology factors and personality disorders traits.

**Results**

*Structure factor analysis*

In the first phase, on the five-factor solution from confirmatory analysis of 30-NEO-FFI (Körner et al., 2008), a principal component factor analysis with Varimax rotation and n-factors = 5 were made including all 30 items. Loadings, communalities, eigenvalues and variance for five-factor structure are summarized in table 1. All loadings for the five factors were higher than |.40| except for item 44 (“I always try to be considerate and sensitive”) in A trait.

In the second phase, each of the five dimensions from 30-NEO-FFI were modeled, individually, as single-factor models to the CFAs. Table 2 shows model fit statistics associated with each dimension obtained from the total sample and from both men and women samples. With no change in the original items for each trait, all chi-square likelihood ratios were significant, indicating poor model fit. However, for N, E, O, and C factors the root-mean-square-error of approximation indices (RMSEA; Steiger, 1990) were higher than .05, and the comparative fit indices (CFI; Bentler, 1990) were higher than .90. Both RMSEAs and CFIs indices showed a good fit for the four dimensions, for both total sample and each one gender samples. For A, only men sample had a good fit (RMSEA = .055), but CFIs were adequate for the total sample and both gender samples.

Table 1 also shows loadings from the five separate CFAs, having similarities between both PAF and CFA structures. Reliability indices for the short version were smaller (ranged from .63 for A to .77 for N) than for the original version (ranged from .84 for A to .91 to N).
Normative Trait scores in both versions

Trait scores were average in both versions. Total score for each trait in the original version was divided by 48, and the total score for each trait in the short version was divided by 6. Test-retest analyses were made for the five traits. Table 3 shows means and standard deviations for all traits in both versions, and r and t indices. All factors had rank order stability, ranging from .75 for A to .85 for N. Nevertheless, the mean levels for participants were higher for N and O, and smaller for E, A, and C in the original ver-
sion than in the short version. The effect size was medium for E, A and C, and small for N.

**Differential validity**

In both versions there were gender differences for N, A and C, with women scoring higher than men, however, the effect size were small (Cohen, 1969). No sex differences were found for E. In case of O, women scored higher than men only in the original version (see table 4).

**Concurrent validity**

Concurrent validity of both versions was established by examining correlations between the five traits from both the original and the short versions of NEO-PI-R, and the nine scales of SCL-90-R and the nine personality disorders

Table 2
**Fit indices of Confirmatory Factor Analysis**

<table>
<thead>
<tr>
<th></th>
<th>Absolute fit</th>
<th>Incremental fit</th>
<th>Parsimony</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X^2$</td>
<td>RMSEA</td>
<td>GFI</td>
</tr>
<tr>
<td>Total</td>
<td>152.359***</td>
<td>.087</td>
<td>.974</td>
</tr>
<tr>
<td>N</td>
<td>84.773***</td>
<td>.106</td>
<td>.959</td>
</tr>
<tr>
<td>Men</td>
<td>92.023***</td>
<td>.083</td>
<td>.976</td>
</tr>
<tr>
<td>Women</td>
<td>196.193***</td>
<td>.100</td>
<td>.968</td>
</tr>
<tr>
<td>E</td>
<td>59.737***</td>
<td>.087</td>
<td>.974</td>
</tr>
<tr>
<td>Women</td>
<td>144.769***</td>
<td>.106</td>
<td>.964</td>
</tr>
<tr>
<td>O</td>
<td>381.915***</td>
<td>.141</td>
<td>.940</td>
</tr>
<tr>
<td>Men</td>
<td>174.448***</td>
<td>.157</td>
<td>.926</td>
</tr>
<tr>
<td>Women</td>
<td>197.206***</td>
<td>.124</td>
<td>.951</td>
</tr>
<tr>
<td>A</td>
<td>37.623***</td>
<td>.039</td>
<td>.994</td>
</tr>
<tr>
<td>Men</td>
<td>29.410***</td>
<td>.055</td>
<td>.987</td>
</tr>
<tr>
<td>Women</td>
<td>17.779*</td>
<td>.027</td>
<td>.996</td>
</tr>
<tr>
<td>C</td>
<td>144.609***</td>
<td>.085</td>
<td>.975</td>
</tr>
<tr>
<td>Men</td>
<td>62.603***</td>
<td>.090</td>
<td>.970</td>
</tr>
<tr>
<td>Women</td>
<td>100.549***</td>
<td>.087</td>
<td>.974</td>
</tr>
</tbody>
</table>

* = $p < .05$; ** = $p < .01$; *** = $p < .001$

Table 3
**Differences between original and short versions**

<table>
<thead>
<tr>
<th></th>
<th>Original version</th>
<th>Short version</th>
<th>$r$</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1.93</td>
<td>.47</td>
<td>1.75</td>
<td>.83</td>
<td>.85***</td>
</tr>
<tr>
<td>E</td>
<td>2.36</td>
<td>.39</td>
<td>2.60</td>
<td>.62</td>
<td>.83***</td>
</tr>
<tr>
<td>O</td>
<td>2.49</td>
<td>.38</td>
<td>2.43</td>
<td>.77</td>
<td>.79***</td>
</tr>
<tr>
<td>A</td>
<td>2.49</td>
<td>.34</td>
<td>2.78</td>
<td>.62</td>
<td>.75***</td>
</tr>
<tr>
<td>C</td>
<td>2.41</td>
<td>.44</td>
<td>2.70</td>
<td>.69</td>
<td>.83***</td>
</tr>
</tbody>
</table>

Note: $r = \text{Pearson's } r$; $t = \text{test-retest Student's } t$; ** = $p < .01$; *** = $p < .001$
of IPDE (Table 5). Intercorrelation patterns between both personality versions and symptomatology were similar. The highest correlations were found for N (in both versions) and all SCL-90-R scales, mainly with Obsessive-Compulsive, Interpersonal Sensitivity, Depression and Anxiety (higher than .45), and the lowest correlations were found between O and all SCL-90 scales.

Table 5 also shows the intercorrelations between both NEO-PI-R versions and dimensional scores for IPDE personality disorders. Again, relation patterns are very similar for both versions. There were positive and significant relations, higher than .30, between N and Paranoid, Borderline, Anxious and Dependent personality disorders; negative relations between E and Schizoid and Anxious personality disorders; negative relations between A and three personality disorders (Paranoid, Disocial, and Impulsive); a single positive relation between C and Anankastic personality disorder; and no relations between O and IPDE factors.

**Discussion**

The findings of five CFAs supported the invariance of factor structure of NEO-FFI-30-SF. Although the chi-squared indices were significant, indicating a poor fit of the model, these findings can be explained because chi-square is overly sensitive to sample size (Campbell-Sill & Brown, 2005). Indeed, the RMSEA indices were adequate for 4 of the 5 factors (except for A), supporting the model fit.

The internal consistency indices for NEO-FFI-30-SF were adequate except for A, showing a small reduction compared to the NEO-PI-R reliability. Considering that alpha depends on the number of items that compose the factor, the reduction to one eighth of the total number of items (from 48 to 6) did not imply a significant decrease in the psychometric goodness of fit of the short version.
Intercorrelations between both NEO versions indicates that N is the most pervasive personality trait as Costa & McCrae (1988) proposed. Similarly to Schmitz et al. (2001), symptomatology from SCL-90-R was mainly related to N, with almost identical $r$-scores. Further more, as it was shown previously (García et al., 2022), the current results show a high convergent validity between the five-factor model and a measure of personality disorders.

The study’s main limitation was that the gains in convenience are met at the expense of loss of information concerning NEO-PI-R facets. On the other hand, the use of samples of convenience, mostly college age, might affect the generalizability of the results to the population. Given that the mean levels for participants in the NEO-FFI-30-SF five factors are different from those obtained in the NEO-PI-R, with medium effect size for E, A and C, despite maintaining stability in the range, would make it necessary to recalculate the normative values with a representative sample.

In sum, the NEO-FFI-30-SF is a reliable personality assessment tool that shows the primary dimensions of personality proposed in the Five Factor Model in Spanish participants.
References


