Psicothema

Psicothema 2013, Vol. 25, No. 1, 100-106
doi: 10.7334/psicothema2011.411

ISSN 0214 - 9915 CODEN PSOTEG
Copyright © 2013 Psicothema
www.psicothema.com

Development and validation of the Overall Personality Assessment Scale (OPERAS)

Andreu Vigil-Colet, Fabia Morales-Vives, Elisa Camps, Jordi Tous and Urbano Lorenzo-Seva
Universitat Rovira i Virgili

Abstract

Background: The aim of the present research was to develop a short measure for the five-factor model personality traits, which allows scores free of social desirability and acquiescence effects to be obtained. **Method:** Scales were created using recently developed methods to control response bias effects in a sample of 3,838 participants from a wide age range. The scales were answered in person or on-line. **Results:** Exploratory and confirmatory factor analysis showed the expected five factor structure. Furthermore, the scales showed good psychometric properties in that they had good reliability, temporal stability and convergence with the widely used Big-Five measures. **Conclusions:** The test developed in the paper presents acceptable psychometric properties, and it is suitable for individuals up from 13 years old. Because the method used to control response bias means that scales' scoring is based upon factorial scores instead of raw scores, we have also developed an internet public application that can be used to obtain these scores.

Keywords: Five-factor model, personality assessment, social desirability, acquiescence.

Resumen

Desarrollo y validación de las Escalas de Evaluación Global de la Personalidad (OPERAS). **Antecedentes:** La presente investigación se centra en el desarrollo de una medida breve de los rasgos de personalidad del modelo de los cinco grandes factores que permita la obtención de puntuaciones libres de los sesgos de respuesta más habituales. **Método:** Para ello la escala se construyó utilizando técnicas desarrolladas recientemente para el control de la deseabilidad social y la aquiescencia y se aplicó a una muestra de 3.838 participantes en un amplio rango de edad. Los participantes respondieron a la misma presencial o virtualmente. **Resultados:** Tanto el análisis exploratorio como el confirmatorio mostraron la estructura esperada de cinco factores. Además, las diferentes subescalas mostraron una buena consistencia interna, estabilidad temporal y validez convergente con otras medidas ampliamente utilizadas en la evaluación de la personalidad. **Conclusiones:** El test desarrollado presenta propiedades psicométricas aceptables, y es adecuado para individuos a partir de 13 años. Debido a que el método utilizado para controlar los sesgos de respuesta precisa de la utilización de puntuaciones factoriales se ha desarrollado una aplicación pública de internet que puede ser utilizada para obtener dichas puntuaciones.

Palabras clave: Modelo de los cinco grandes, evaluación de la personalidad, deseabilidad social, aquiescencia.

In recent decades the study of personality has been highly focused on the lexical Big Five Model and on its related questionnaire approach represented by the Five Factor Model (Goldberg, 1981, 1990; Costa & McCrae, 1992). This model has considerable support and has been successfully applied to numerous areas of psychology, making it the most widely used model of personality (John, Naumann, & Soto, 2008).

Consequently, several instruments have been developed to measure its dimensions. Among the most widely used is the NEO Personality Inventory Revised (NEO-PI-R), which includes not only the big five domains (Extraversion, Agreeableness, Emotional Stability, Conscientiousness and Openness to Experience) but also six specific facets within each dimension (Costa & McCrae, 1992). At 240 items, this instrument is quite long, and so shorter instruments have been developed, such as the 60-item alternative NEO - FFI (Costa & McCrae, 1992). These are especially interesting when time is a key issue in test administration. In the lexical approach there are also short measures such as the 44-item Big Five Inventory (Benet-Martínez & John, 1998).

As with other measures in the personality domain, these are not free of response bias. The two main response biases are acquiescence (AC) and social desirability (SD) (see Paulhus, 1991), and, although in a well-designed questionnaire the prime determinant of the response is the content dimension that the test intends to measure, response biases are expected to influence individuals' scores to a certain degree (Ferrando, Lorenzo-Seva, & Chico, 2009).

Several studies have shown that SD and AC are substantially related to different self-reported measures within the Five Factor Model (FFM) (Furnham, 1997; Holden & Passey, 2010; Kurtz, Tarquini, & Iobst, 2008; Rammstedt, Goldberg, & Borg, 2010;

Received: December 14, 2011 • Accepted: March 28, 2012
Corresponding author: Andreu Vigil-Colet
Centre de Recerca en Avaluació i Mesura de la Conducta
Universitat Rovira i Virgili
43007 Tarragona (Spain)
e-mail: andreu.vigil@urv.cat

Soto, John, Gosling, & Potter, 2011). Although there is no consensus about how these response biases may affect the criterion and predictive validity of questionnaires, they do have effects at two levels at least. First, SD may affect estimates of individual's scores, which may be critical when psychologists have to take decisions on the basis of them. Second, the psychometric properties of personality questionnaires and their factorial structure seem to be affected by different factors such as age, cognitive complexity or educational level, and these effects seem to be caused by differential acquiescence effects (Soto et al., 2008; 2011). For instance, Ramsteed et al. (2010) have shown that the factorial structure of the FFM only holds across educational levels when acquiescence effects are removed, and Soto et al. (2008) described a similar result for age levels.

Traditionally, SD effects have been controlled by administering an SD scale together with scales designed to measure the content of interest. The SD scale is then used to remove individuals with high scores in SD. Nevertheless this method has some limitations. First, removing participants with high scores in SD does not guarantee that the scores of the others are free of SD. Second, if SD is related to the content that is being measured, then by removing individuals with high SD scores we may also be removing individuals with high content scores. Third, if the psychologist is interested in an individual's score, this method cannot give an individual SD-free score for the content of interest.

To correct these deficiencies, Ferrando (Ferrando, 2005; Ferrando, Lorenzo-Seva, & Chico, 2009), proposed a procedure that can be used to simultaneously control response bias caused by SD and AC. It consists of the following three successive steps:

Step 1. The first step consists of identifying a factor related to SD. The four items related to SD are taken as *markers* of SD. The inter-marker correlation matrix is analyzed using factor analysis. The corresponding loading values are used to compute the loading values of the content items on the SD factor using the *Instrumental Variables Technique* (Hägglund, 1982). Once the factor related to SD is available, the variance explained by this factor is removed.

Step 2. The second step consists of identifying a factor related to AC. The residual inter-item correlation matrix is now analyzed and the variance due to acquiescent responding is removed from the content items. As the scales in OPERAS are only partially balanced, the procedure proposed by Lorenzo-Seva and Ferrando (2009) was used. First, the number of expected content factors plus an additional factor are retained. Second, a first general principal component is computed and taken as an estimate of the loading value of each item on the AC factor. Third, this set of estimates (one estimate per item) is used as a target in a congruent rotation to compute the factor loading value of each content item on the AC factor. Once the factor related to AC is available, the variance explained by this factor is removed.

Step 3. The third step consists of identifying the factors related to the measured traits in the residual matrix.

The application of this procedure at the item calibration level provides three loading estimates for each item: a loading on the content factor that the test intends to measure, and two loadings on two orthogonal factors identified as SD and AC, which allows individuals' scores to be obtained free of response bias effects. A full application of the method in the development of a new questionnaire may be found in Morales-Vives, Camps & Lorenzo-Seva (in press).

Our aim was to develop a questionnaire that incorporates this procedure for controlling response bias in order to assess personality in the context of the FFM. Although there are many questionnaires in the FFM framework with good overall psychometric properties, the development of a new one based upon the procedure described above will have two advantages. First, it will be possible to obtain more accurate estimations of these traits because they are free from response bias. Second, the factorial structure of the test will be highly stable across ages and educational levels because the analysis is performed after removing the SD and AQ effects.

Furthermore, given that the FFM has been shown to exist across age groups (Digman & Takemoto-Chock, 1981), we wanted the scale to be suitable for all individuals from adolescence to old age, and short enough to be easily administered both individually and in groups.

Method

Development of the Overall Personality Assessment Scale (OPERAS)

We wanted the OPERAS to assess five uncorrelated traits: Extraversion (EX), Emotional Stability (ES), Conscientiousness (CO), Agreeableness (AG), and Openness to Experience (OE). Taking into account the characteristics of these traits, the authors wrote 60 items, which were rated by 5 judges with experience in developing personality tests. The 50 items with the best ratings were used in a pilot study of 258 undergraduate students and those items with loadings lower than .30 or with complex loadings (greater than .30 in more than one factor) were removed. Finally, the seven items with the highest loadings on each factor were used to create the scales. Each item is a sentence that describes typical situations that may be experienced by individuals. The participant must indicate the level of agreement with the sentence by using a five-point scale that goes from "fully disagree" (1) to "fully agree" (5).

Four items related to SD were included in the test, and 15 content items were worded in the opposite direction to the other items to allow AC control. An additional item was included as the first item of the scale. The aim was to have a dummy item that could be used as a training item when administrating the test via computers (see, for example, Ferrando & Lorenzo-Seva, 2005). The Spanish items and the English translations of the inventory can be obtained at http://psico.fcep.urv.cat/questionaris/operas40/

Participants

A total of 3,838 individuals participated in the study (51% women). The participants were between 13 and 95 years old (mean 29, s.d. 14.3). The distribution of ages and sex is summarized in Table 1. The academic level of the participants was elemental (15%), medium (43%), and higher (41%). The participants described themselves as students (46%), employed (36%), or unemployed (18%). Some of the participants ($N = 128$) were tested a month later in order to assess the test/retest reliability. Another set of participants ($N = 193$) were administered other personality tests to assess convergent validity. These participants were university students between 18 and 43 years old (mean 20.4, s.d. 3.8) in the

first case and between 18 and 49 years old (mean 21.4, s.d. 4.03) in the second.

Procedure

Different strategies were used to collect the sample: the test was administered to (1) groups of adolescents in their classroom or during university open day activities, (2) students at university in their classrooms; (3) the test was advertised on Facebook as an open personality test that offered a short general personality description; (4) the test was used by different human search and selection companies; (5) participants during academic university activities aimed at graduates/adults; and (6) elderly people in their nursing homes. This last set of participants was also tested to exclude people with dementias. The test was administered via the Internet with an application that has already been tested in an applied study (Ferrando & Lorenzo-Seva, 2005). 46% of the questionnaires were collected in this way. A total of 93,9% of the tests administered by internet were answered by Spanish people and 6,1% were answered by people from South and North America. The participants were asked for no information that could have identified them, thus guaranteeing their anonymity.

Instruments

In addition to the OPERAS, two widely used questionnaires based on the Big-Five model of personality, the Big Five Inventory (BFI) and the Five Factors Personality Inventory (FFPI) were used to assess convergent validity. We administered the Spanish versions of these questionnaires (Benet-Martínez & John, 1998; Rodríguez-Fornells, Lorenzo-Seva, & Andrés-Pueyo, 2001).

Data analysis

The sample of 3,838 participants was randomly split in two halves. The first sample was used to conduct an Exploratory Factor Analysis (EFA). The second sample was used to conduct a Confirmatory Factor Analysis (CFA). Both analyses led to the same conclusions, and so the overall sample was used to obtain the factorial weights needed to compute scores. The scale analyses were also computed for the overall sample. EFA and CFA were performed using MATLAB, FACTOR 6.01 (Lorenzo-Seva & Ferrando, 2006), and LISREL 8.5 (Joreskog & Sorbom, 2001) respectively.

Table 1
Percentage of participants tabulated for sex and age

Age	Women	Men
< 21	23.00%	12.78%
21 - 30	14.03%	15.70%
31 - 40	7.07%	12.04%
41 - 50	2.57%	5.00%
51 - 60	1.30%	1.48%
61 - 70	1.32%	0.85%
71 - 80	1.11%	0.69%
> 80	0.66%	0.40%

Results

We computed the polychoric correlation matrix between 39 items from the OPERAS (the first item was excluded from the analysis). The KMO index value was .83. Kaiser and Rice (1974) suggested that any value in the 0.80s was 'meritorious', so the KMO value suggested that the correlation matrix was well suited for factor analysis.

We computed the Parallel Analysis (PA, Horn, 1965) as proposed by Lattin, Carroll and Green (2003), which indicated that there were seven dimensions underlying the data. PA is based on comparing each eigenvalue to random eigenvalues: the aim is to retain only those factors that are related to an amount of variance larger than the amount of variance of random factors. This result was consistent with the five dimensions related to content scales and the two dimensions related to response styles.

We applied the procedure explained above to determine the two response style factors. Five content factors were retained using Minimum Rank Factor Analysis (Ten Berge & Kiers, 1991). To determine the loading factors related to the five content factors, we computed a Varimax rotation. To assess the fit of the rotated loading matrix, we computed the congruence index (Tucker, 1951) between the rotated loading matrix and the *ideal* loading matrix. The congruence values ranged between .85 and .93. As the coefficients were above the threshold of .85, there was a fair factor similarity between the rotated and the ideal loading matrixes (Lorenzo-Seva & Ten Berge, 2006). Finally, we computed Bentler's simplicity (S) index (1977) and the Loading Simplicity (LS) index (Lorenzo-Seva, 2003), which gave values of .88 and .41, respectively. We should emphasize that the values of these indices indicated that our factor solution was not a *truly simple factor solution*. In fact, in the context of the Big Five Model, some items are expected to be to some extent complex items; that is, some items are expected to have a main salient loading value, and at least one secondary loading value that is not as high as the main one, but that is still significant (see for example De Raad 2000, pages 72-73).

In order to study the replicability of the factor structure obtained in the first sample, a CFA was carried out on the second sample. First the variance due to SD and AC was partialized following the procedure proposed by Ferrando et al. (2009). Unweighted least square estimates were computed from the residual covariance. It was proposed that the model should retain five uncorrelated factors, as the exploratory factor analysis explained above suggested. Ferrando & Lorenzo-Seva (2000) pointed out that when personality questionnaires obtained by EFA are tested using CFA, the model proposed is usually rejected, although a series of different exploratory studies have previously replicated the same factorial structure, especially in multidimensional questionnaires. In these cases they propose that semi-restricted models are more appropriate for testing the model fit in CFA. Taking this into account, we applied this analysis by selecting five marker items. To select these items, we selected the simplest items from the previous exploratory factor analysis.

Although there is a lack of agreement about what the CFA cut-off values should be when assessing model adjustment, there is certain consensus that values equal or greater to .90 for the Comparative Fit Index (CFI) and the Goodness of Fit Index are an acceptable fit, while values under .08 in the Root Mean Square Error of Approximation (RMSEA) are acceptable and below .05 are excellent (Bentler, 1990; Hu y Bentler, 1999). The values obtained for these indices in our study were CFI = .97, GFI = .98,

and RMSEA= 0.034. Therefore, it was concluded that the data show a good fit to the proposed model.

Given that both exploratory and confirmatory studies led to similar conclusions, we used the whole sample (N= 3,838) to estimate the factor loadings, and the weights on estimate factor scores. The aim was to use the largest possible sample in order to obtain the best possible estimates. Table 2 shows the loading values after rotation. The loading values for the content factor show that the items were related to the corresponding expected scale.

As the table shows, most of the items related to ES, CO and AG loaded on the SD scale, whereas the other scales were completely free of SD. Furthermore, most of the items loaded on the AC scale.

Table 2
Loading matrix obtained in the final factor analysis. In bold face expected salient loadings

Items	Control Scales		Content Scales				
	SD	AC	EX	ES	CO	AG	OE
5. Siempre mantengo mi palabra	**-0.34**						
11. Alguna vez he cogido algo que no era mío	**0.78**						
19. Alguna vez he dicho algo malo de alguien	**0.61**						
26. Alguna vez me he aprovechado de alguien	**0.70**						
2. Soy el alma de la fiesta	0.06	0.09	**0.63**	0.16	-0.01	0.00	-0.08
8. Me desenvuelvo bien en situaciones sociales	-0.10	0.12	**0.61**	0.31	0.19	0.10	0.09
14. Hablo poco	-0.05	0.13	**-0.65**	-0.03	-0.11	-0.07	-0.06
20. Hago amigos con facilidad	-0.08	0.18	**0.64**	0.30	0.12	0.17	-0.03
25. Prefiero que otros sean el centro de atención	0.00	0.22	**-0.65**	0.02	0.09	0.12	0.00
31. Permanezco en segundo plano	0.00	0.27	**-0.68**	-0.09	0.01	0.03	-0.02
36. Sé cautivar a la gente	0.00	0.17	**0.55**	0.33	0.14	0.01	0.05
3. Me siento cómodo conmigo mismo	-0.13	0.00	0.18	**0.69**	0.18	-0.02	-0.04
9. A menudo tengo el ánimo por el suelo	0.17	0.48	-0.13	**-0.74**	-0.02	-0.11	-0.06
15. A menudo me siento triste	0.19	0.39	-0.21	**-0.60**	-0.11	-0.17	-0.03
21. Es difícil que las cosas me preocupen	0.01	0.03	0.02	**0.42**	-0.27	-0.05	-0.17
27. Me dejo llevar por el pánico con facilidad	0.18	0.25	-0.15	**-0.46**	-0.14	-0.12	-0.08
32. Cambio de humor a menudo	0.30	0.41	-0.06	**-0.54**	-0.04	-0.30	-0.02
37. Me desagrado	0.15	0.36	-0.15	**-0.71**	-0.14	-0.04	-0.07
4. Siempre estoy dispuesto a asumir responsabilidades	-0.24	0.12	0.17	0.29	**0.46**	0.05	0.05
10. Evito mis obligaciones	0.33	0.22	-0.09	-0.10	**-0.57**	-0.11	-0.16
16. Dejo las cosas a medias	0.35	0.26	-0.09	-0.15	**-0.60**	-0.06	-0.04
22. Dejo mis cosas desordenadas	0.34	0.20	-0.03	-0.10	**-0.49**	-0.08	-0.01
28. Soy perfeccionista	-0.05	0.17	0.04	0.11	**0.42**	-0.19	0.24
33. Pierdo el tiempo	0.39	0.23	-0.13	-0.24	**-0.46**	-0.13	-0.04
38. Cuando hago planes los mantengo	-0.21	0.18	0.08	0.24	**0.34**	0.10	0.00
6. Suelo hablar bien de los demás	-0.30	0.25	0.06	0.26	0.03	**0.42**	0.02
12. Respeto a los demás	-0.32	0.26	0.00	0.24	0.27	**0.48**	0.18
17. Creo que los demás tienen buenas intenciones	-0.10	0.13	0.06	0.15	-0.02	**0.39**	0.06
23. Soy muy crítico con los demás	0.28	0.15	-0.01	-0.05	0.01	**-0.49**	0.09
29. A menudo soy desagradable con otras personas	0.34	0.20	-0.12	-0.20	-0.15	**-0.46**	-0.14
34. Acepto a la gente tal y como es	-0.19	0.21	0.05	0.22	0.08	**0.48**	0.04
39. Cuando alguien me la juega, se la devuelvo	0.34	0.15	0.02	0.01	-0.01	**-0.34**	-0.24
7. El arte me parece aburrido	0.08	0.05	0.00	0.05	0.00	-0.02	**-0.67**
13. Creo en la importancia de formarse culturalmente	-0.08	0.26	-0.04	0.19	0.22	0.05	**0.56**
18. Evito las discusiones filosóficas	-0.08	0.13	-0.09	0.00	0.08	0.09	**-0.44**
24. Me gusta visitar museos	-0.14	0.07	-0.08	0.04	-0.02	-0.06	**0.57**
30. Me gusta visitar sitios nuevos	0.00	0.28	0.15	0.16	0.14	0.18	**0.42**
35. Siento curiosidad por el mundo que me rodea	0.02	0.26	0.13	0.18	0.16	0.16	**0.50**
40. El teatro me parece poco interesante	0.11	0.04	-0.03	0.06	-0.06	-0.06	**-0.65**
% Variance accounted	12.47	7.25	12.59	14.38	8.49	6.96	10.46

SD accounted for 12.47% of the variance, whereas AC accounted for 7.25% of the variance. That is to say that the response styles factors explained as much variance as each of the content factors.

As has been mentioned, individuals' scores for the OPERAS must be obtained using factor score estimates. We computed factor scores following the procedure proposed by Ten Berge *et al*. (1999). The factor weights used to compute these factor scores are available on request from the authors.

We computed the reliability estimates on the basis of the factor scores for the scales. In addition, test-retest reliabilities were computed. The reliabilities are shown in Table 3. While EX, ES, and EO showed acceptable reliabilities and temporal stability for the factor scores, CO and AG were less reliable scales. However, when we computed 90% Bootstrap confident intervals for Test-retest reliabilities, CO and AG did not significantly differ from the threshold of .80. We should point out that, because we designed the subscales to be just seven items long, we did not expected the reliabilities of subscales to be high. However, we still preferred to have a short instrument rather than a long instrument that was difficult to administrate to certain individuals in the population (for example, adolescents or old people).

The Kolmogorov-Smirnov test showed that none of the score distributions in the population differed significantly from a normal distribution. Psychological reports for individuals can therefore be based on normalized percentiles.

Table 4 shows product moment correlations between the OPERAS, BFI and FFPI scales. As can be seen, all the OPERAS scales showed good convergent validity, having the highest correlation of each scale with the equivalent scale of the BFI and the FFPI.

We also computed factor scores for the OPERAS scales without removing the response bias effects. It is worth mentioning that the validity coefficients between the OPERAS, BFI and FFPI scales were greater for the scores without response bias corrections, especially for the scales most affected by DS, so traditional validity coefficients may be overestimated due to response bias effects. Nevertheless, the difference between them was only significant for CO ($z= 3.34$ $p<0.01$ and $z= 2.1$ $p<0.05$ for BFI and FFPI respectively).

Discussion

The data reported below show that the OPERAS is a quick system with good reliability and validity for personality assessment in the FFM model domain. Nevertheless if these were the only features of OPERAS, it would merely be another test to add to the different existing measures within the FFM model framework. As such it would be of limited value given that there are already many measures with different administration times that have proved their value for measuring personality.

Table 3
Descriptive statistics for scales

	Correlation between items and their corresponding scale		**Reliability of factor scores**	**Test-retest reliability**	
	Minimum	**Maximum**		**Point estimate**	**90% confidence interval**
EX	0.13	0.54	0.86	0.70	(0.53 - 0.84)
ES	0.11	0.57	0.86	0.70	(0.61 - 0.77)
CO	0.07	0.46	0.77	0.75	(0.68 - 0.81)
AG	0.16	0.42	0.71	0.73	(0.65 - 0.80)
OE	0.17	0.50	0.81	0.79	(0.72 - 0.84)

Table 4
Correlations between OPERAS, BFI and FFPI

		EX	**AG**	**CO**	**ES**	**OE**
BFI	EX	**.792 (.794)**				
	AG	*.182 (.175)*	**.608 (.690)**			
	CO	.015 (.017)	.024 (.102)	**.665 (.816)**		
	NE	-.040 (-.029)	**-.274 (-.263)**	-.069 *(-.158)*	**-.625 (-.689)**	
	OE	**.355 (.347)**	.114 (0.099)	**.218** *(.142)*	**.226** *(.170)*	**.603 (.611)**
FFPI	EX	**.644 (.635)**				
	AG	**-.226 (-.221)**	**.413 (.491)**			
	CO	**-.207 (-.198)**	.098 **(.205)**	**.543 (.672)**		
	ES	*.147* (.135)	**.340 (.340)**	**.188 (.249)**	**.742 (.803)**	
	AU	**.405 (.395)**	.090 (.084)	**.252 (.250)**	**.416 (458)**	**.449 (.453)**

Note: EX: extraversion; AG: agreeableness; CO: conscientiousness; ES: emotional stability; OE: openness to experience; AU: autonomy. Correlations without removing response bias effects are shown in parentheses.
p<0.05 **p<0.01**

The main advantage of the OPERAS when it is compared with these measures is that it is the first to provide scores that are free of the two best known response biases: social desirability and acquiescence. This is no minor point if we take into account the fact that, as the results below show, SD and AC account for as much and, in certain cases, even more variance than some of the factor contents.

As other authors have reported (Holden & Passey, 2010; Ones, Viswesvaran, & Reiss, 1996), the scales most affected by SD were ES, CO and AG, whereas its effects on EX and OE were almost negligible. On the other hand, AC had important effects on all the OPERAS scales. Controlling these effects in order to increase measurement accuracy is desirable for several reasons.

First, in many situations psychologists may be interested in selecting or excluding individuals with extreme scores on some scales (i.e. during a selection process). If the trait is highly undesirable or desirable, the effects of SD on individuals' scores may imply that some of the individuals chosen do not meet the established criteria. Second, although there is a lack of agreement about the possible effects of SD on validity, our data seem to show that when both SD and AC are taken into account, psychologists can overestimate the validity of the measures, especially those that are most affected by SD, such as CO. Finally, AC seems to have differential effects linked to variables such as age, educational level, etc, which may mean that the precision of the measurement instruments in these groups is lower if these effects are not removed. Furthermore, depending on which of these variables is measured, the response bias seems to affect not only the accuracy but also the factorial structure of the model. This implies that the FFM lacks factorial invariance across, for instance, educational levels if AC is not controlled (Ramsteed et al., 2010). We think that these reasons demonstrate the usefulness of tests such as the OPERAS, especially when response bias may be a critical issue for the psychologist when applying the test to a given population or situation.

It should be mentioned that psychologists typically like to compute scores of psychological tests as raw additions of individuals' answers to the items (raw scores). From this point of view, a drawback of the OPERAS is that the scores to be interpreted must be factor scores (not raw scores). To compute factor scores, individuals' answers to items must be standardized using the mean and standard deviations. The standardized responses must be added as a weighted addition (the weights to be applied are shown in Table 2). Finally, the total addition should be transformed from typical scores into T scores (i.e., mean 50 and standard deviation 10). In addition, normalized percentiles should to be computed in order to complement the proper psychological report. Although this is not a complex procedure, it is still not straightforward for applied psychologists. To solve this drawback, we have developed a public internet application (http://psico.fcep.urv.cat/questionaris/operas40/) that applied psychologists can use to obtain scores SD and AQ free, normalized percentiles, and a short written report describing an individual's score.

Acknowledgments

The research was supported by a grant from the Spanish Ministry of Economy and Competitivity (PSI2011-22683).

References

Benet-Martínez, V., & John, O.P. (1998). *Los Cinco Grandes* across cultures and ethnic groups: Multitrait method analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729-750.

Bentler, P.M. (1977). Factor simplicity index and transformations. *Psychometrika*, *59*, 567-579.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238-246.

Costa, P., & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

De Raad, B. (2000) *The big five personality factors*. Göttingen. Hoegrefe & Hubert.

Digman, J.M., & Takemoto-Chock, N.K. (1981). Factors in the natural language of personality: Reanalysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research, 16,* 149-160.

Ferrando, P.J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behavioral Research, 40,* 331-349.

Ferrando, P.J., & Lorenzo, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicologica, 21,* 301-323.

Ferrando, P.J., & Lorenzo-Seva, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and Internet-administered questionnaires. *Psychological Methods*, *10*, 193-205.

Ferrando, P.J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling*, *16*, 364-381.

Furnham, A.F. (1997). Knowing and faking one's five factor personality score. *Journal of Personality Assessment, 69,* 229-243.

Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-160). Beverly Hills, CA: Sage Publications.

Goldberg, L.R. (1990). An alternative "description of personality": The Big Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216-1229.

Hägglund, G. (1982). Factor analysis by instrumental variable methods. *Psychometrika*, *47*, 209-222.

Holden, R.R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49,* 446-450.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

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

John, P.O., Naumann, L.P., & Soto, C. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O.P. John, R.W. Robins, & L.A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3rd edition, pp. 114-158). New York. Guilford.

Joreskog, K., & Sorbom, D. (2001). *LISREL 8.5: User's Reference Guide*. Chicago: Scientific Software International.

Kaiser, H.F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*, 111-117.

Kurtz, J.E., Tarquini, S.J., & Iobst, E.A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences, 45,* 22-27.

Lattin, J., Carroll, D.J., & Green, P.E. (2003). *Analyzing multivariate data* (pp. 114-116). Duxbury Press.

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, *68*, 49-60.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the Exploratory Factor Analysis model. *Behavior Research Methods, Instruments, & Computers*, *38,* 88-91.

Lorenzo-Seva, U., & Ferrando, P.J. (2009). Acquiescent responding in partially balanced multidimensional scales. *British Journal of Mathematical and Statistical Psychology*, *62*, 319-326.

Lorenzo-Seva, U., & Ten Berge, J.M.F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57-64.

Morales-Vives, F., Camps, E., & Lorenzo-Seva, U. (in press). Development and validation of the Psychological Maturity Assessment Scale (PSYMAS). *European Journal of Psychological Assessment*.

Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660-679.

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA. Academic Press.

Rammstedt, B., Goldberg, L.R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44,* 53-61.

Rodríguez-Fornells, A., Lorenzo-Seva, U., & Andrés-Pueyo, A. (2001). Psychometric properties of the Spanish adaptation of the five factor personality inventory. *European Journal of Psychological Assessment, 17,* 145-153.

Soto, C.J., John, O.P., Gosling, S.D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*, 718-737.

Soto, C.J., John, O.P., Gosling, S.D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100,* 330-348.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309-315.

Ten Berge, J.M.F., Krijnen, W., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications*, *289*, 311-318.

Tucker, L.R. (1951). *A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984)*. Department of the Army, Washington D.C., Washington D.C.