

How response biases affect the factor structure of Big Five personality questionnaires

Fabia Morales-Vives*, Urbano Lorenzo-Seva and Andreu Vigil-Colet

Research Center for Behavior Assessment (CRAMC). Psychology Department, Universitat Rovira i Virgili, Tarragona (Spain).

Título: Cómo afectan los sesgos de respuesta a la estructura factorial de los tests basados en el modelo de los Cinco Grandes factores de personalidad.

Resumen: Diversos estudios muestran que los tests de personalidad pueden verse afectados por los sesgos de respuesta, lo que puede tener consecuencias en su estructura factorial, especialmente en muestras con bajo nivel educativo o adolescentes. El objetivo del presente estudio consiste en determinar cuál es el efecto de la deseabilidad social y la aquiescencia en la estructura factorial de tres cuestionarios basados en el modelo de los Cinco Grandes factores de personalidad: Big Five Inventory, Five Factor Personality Inventory y Overall Personality Assessment Scale. Los datos se analizaron a partir de un nuevo método que elimina los efectos de la deseabilidad social y la aquiescencia de la matriz de correlaciones inter-item utilizada en el análisis factorial. La muestra está compuesta por 392 estudiantes universitarios, por lo que no incluye individuos con bajo nivel educativo, niños o adolescentes. Los resultados indican que, incluso en muestras sin individuos con bajo nivel educativo, controlar la deseabilidad social y la aquiescencia da lugar a una estructura factorial más simple y congruente con la solución teórica esperada a partir del modelo de los Cinco Grandes. Además, los resultados sugieren que en los tests basados en este modelo, este efecto puede ser debido especialmente a la aquiescencia.

Palabras clave: Modelo de los Cinco Grandes; personalidad; deseabilidad social; aquiescencia; sesgos de respuesta.

Abstract: Several studies have shown that personality self-reports may be affected by response biases, and that this may have consequences on their factor structure, especially in samples with little education or in adolescents. The current study aims to understand the effect of social desirability and acquiescence on the factor structure of three questionnaires based on the Five Factor Model of personality: the Big Five Inventory, the Five Factor Personality Inventory and the Overall Personality Assessment Scale. The data was analysed using a new method that removes the effects of both social desirability and acquiescence from the inter-item correlation matrix used for factor analysis. These effects were assessed in a sample of 392 university students, which contained no individuals with low educational levels, children or adolescents. The results showed that, even in samples with no individuals with low educational levels, controlling for social desirability and acquiescence led to a simpler factor structure that is more congruent with the theoretical solution expected from the five factor model. It also seems that in the domain of inventories based upon the five factor model, this effect may be specially due to acquiescence.

Key words: Five Factor Model; personality; social desirability; acquiescence; response biases.

Introduction

Throughout the 20th century, personality was conceptualized from a variety of theoretical perspectives, which gave rise to a wide range of traits and many different scales by which they were measured. This complexity is evidenced by the fact that some scales measure different things with the same name, while others measure the same thing with different names (John & Srivastava, 1999). For this reason a general taxonomy of personality was needed to be found to facilitate communication among researchers and make possible a global understanding of personality. The Five Factor Model of personality (FFM) has provided this general taxonomy and a common framework within which personality can be described. Unlike other systems of personality, this system is not based on previous theoretical approaches that make it difficult to comprehensively identify facets and features (cf. De Raad & Hendriks, 1997). It is based on the analysis of language and it is known as the psycholexical approach. According to Goldberg (1981), the most significant individual differences in daily transactions with other people will be represented in language, so the analysis of the vocabulary used will enable the individual differences to be identified. Thanks to this procedure, the FFM is more comprehensive than any system previously proposed. The FFM, and the questionnaires derived from it, have had a considerable

impact on psychology in recent decades, and it is the most widely used model of personality nowadays (John, Naumann & Soto, 2008; Ruiz & Jiménez, 2004).

Several questionnaires have been developed to measure the big five personality traits. One of the most popular is the NEO Personality Inventory Revised (NEO-PI-R; Costa & McCrae, 1992) which assesses the following traits: Agreeableness, Extraversion, Emotional Stability, Conscientiousness and Openness to Experience. Each of these traits includes six specific facets. It is a long questionnaire, with 240 items. Another well-known questionnaire is the Five-Factor Personality Inventory (FFPI; Hendriks, Hofstee, & De Raad, 1999) which assesses all the traits included in NEO-PI-R except Openness to experience. In fact, the fifth factor in FFPI is named Autonomy, and it refers to the tendency to take independent decisions without being influenced by the social environment, and to maintain an independent point of view. It is also a long questionnaire, with 100 items, 20 for each of the scales. Shorter instruments have been developed although they do not usually measure specific facets. The advantage of shorter inventories is that they can efficiently assess the big five dimensions when there is no need to assess individual facets, saving time and avoiding fatigue and boredom (Burisch, 1984). One of these questionnaires is the Big Five Inventory (BFI; Benet-Martínez & John, 1998), which consists of 44 items and measures the same traits as NEO-PI-R.

One issue that has been mentioned quite often in relation to the psychometric quality of FFM measures is the effect of response bias. Several authors have pointed out the importance of controlling response biases, particularly social

* Correspondence address [Dirección para correspondencia]:

Fabia Morales Vives. Universitat Rovira i Virgili. Departament de Psicologia.
Campus Sescelades. Carretera de Valls s/n. 43007 Tarragona (Spain). E-mail:
fabia.morales@urv.cat

desirability (SD) and acquiescence (AC) in personality measures within the FFM domain (e.g., Holden & Passey, 2010; Soto, John, Gosling, & Potter, 2011; Soubelet & Salthouse, 2011; Vigil-Colet, Morales-Vives & Lorenzo-Seva, 2013). Some studies have shown that acquiescence and, specially, social desirability increase considerably with age. At the individual level this means that that self-reports in the elderly that do not control for this effect should be viewed with caution and, at the group level, some of the relationships found between age and various personality dimensions may reflect response bias effects and not true variations in the personality dimensions (Dijkstra, Smit, & Comijs, 2001; Soubelet & Salthouse, 2011; Stöber, 2001; Vigil-Colet, Morales-Vives & Lorenzo-Seva, 2013; Vigil-Colet, Lorenzo-Seva & Morales-Vives, 2015).

Acquiescence can also depend on educational levels and cognitive complexity so it is advisable to use questionnaires that correct response biases, so that there are no issues of inequality when individuals with low education or low intellectual competence are assessed (Meisenberg & William, 2008; Soto, John, Gosling & Potter, 2008, 2011).

Although response bias may have an impact on individual scores, this article focuses on how response bias can have an undesirable effect on the factor structure of questionnaires. Several studies have analysed this effect in the case of AC, and the findings may be easily generalized to the case of SD. One problem with this, however, is that individuals show different levels of AC and SD. In the case of AC, this variability in response bias tends to overestimate the covariance between items worded in the same direction, underestimate the covariance between items worded in different directions, and overestimate the correlation between socially desirable or undesirable items (Paulhus & Vazire, 2007; Soto et al., 2011). The resulting distortion in the inter-item covariance matrix has effects on both exploratory (EFA) and confirmatory factor analysis (CFA). In the case of EFA this may result in complex saturations because items designed to measure a particular content share variance due to response bias with items that measure other factors and, in the case of AC, one factor may emerge representing the within-person mean item response, or two factors representing positively worded and negatively worded items (McCrae, Herbst & Costa, 2001; Paulhus & Vazire, 2007; Ten Berge, 1999). Another consequence of response bias in the factor structure of questionnaires is that it may be difficult to replicate the expected factor structure of the FFM if some groups show high levels of response bias because of their educational level or age (i.e. acquiescence) (Rammstedt, Goldberg & Borg, 2010; Rammstedt & Farmer, 2013).

On the other hand, if CFA is used to confirm the proposed structure of the theoretical model underlying the tests, response biases may generate misfit in the model, because of the distortions that are generated in the inter-item covariance matrix. For instance, two items measuring the same content and affected by SD may show a correlation that cannot be totally explained by the latent variable which rep-

resents the content that they are measuring. On the other hand, if the two items measure different contents, their correlation may not be totally explained by the correlation between the two latent variables they are measuring. If different items are impacted by SD and AC the misfit generated by each pair of items is accumulative and if the effects of response biases are not controlled before the model is tested the fit of the model may be worse.

Although the reasons that explain the effects of AC on the factor structure of personality questionnaires are the same in the case of SD, most research on the effects of response bias has focused on AC. This may be because the effects of AC on the inter-item correlation matrix can be controlled (i.e. ipsatization of item responses) but the effects of SD are more difficult to remove.

One procedure that provides an inter-item correlation matrix free not only of AC but also of SD was developed by Ferrando (2005) and Ferrando, Lorenzo-Seva, and Chico (2009). This procedure requires to include in the questionnaire a set of items related to SD, which are taken as markers of SD. Moreover, it also requires that some items of the questionnaire are worded in the opposite direction to the other items, to identify acquiescent response style. The first step is to identify a factor related to SD by using the items that are markers of SD. More specifically, the inter-marker correlation matrix is analysed using factor analysis to obtain the corresponding loading values of each marker in the SD factor. When the SD factor is defined, these loading values are used to compute the loading values of the content items on the SD factor using the *Instrumental Variables Technique* (Häggglund, 1982). The next step is to remove the variance explained by the SD factor, which leads to an inter-item residual correlation matrix with no SD effects. Then, if the questionnaire has items positively and negatively worded items, the method developed by Lorenzo-Seva and Ferrando (2009) can be used. This method is based on the idea that, in a balanced scale, the centroid helps to estimate the overall tendency of individuals to systematically use a particular value of the response scales independently of the worded direction of the items (i.e., to show an acquiescent response style). In an initial step, the first centroid is computed and taken as an estimate of the loading values of items on an underlying acquiescent factor. If the scale is partially balanced, a subset of balanced items is used to compute the first centroid, and then the unbalanced set of items is projected on the first centroid. The variance explained by the first centroid is then removed from the correlation matrix, and the residual correlation matrix is factor analysed to estimate the loading on the content factors. Therefore, this procedure removes the variance due to acquiescent responding, and provides a second inter-item residual correlation matrix free of SD and AC. Finally, this residual matrix can be used as an input matrix in an EFA or a CFA. The authors have recently developed the program Psychological Test Toolbox that makes it easy for researchers and professionals to use these

procedures. This program can be downloaded from the following website:

<http://psico.fcep.urv.es/utilitats/PsychologicalTestToolbox/index.html>

The main advantage of this procedure is that it provides the loadings on the content factor (in this case the big five personality traits) and also the loadings on the SD and AC factors, which makes it possible to obtain individuals' scores free of response bias effects. This method has been successfully used to develop new questionnaires, and has proved to be effective at controlling these response biases, but it also may be applied to questionnaires not initially developed with this method, as long as SD markers are added to them (Morales-Vives, Camps, & Lorenzo-Seva, 2013; Vigil-Colet, Morales-Vives, Camps, Tous, & Lorenzo-Seva, 2013; Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco & Lorenzo-Seva, 2012). The method can also use the inter-item correlation matrix with or without the response bias effect as input, which makes it possible to compare the results of both analyses and assess the effects of controlling response biases on an individual's scores, the factor structures, or the validity of the test.

According to some authors, procedures to control response biases in personality questionnaires may partial out relevant content variance, what would decrease validity coefficients (e.g., McCrae & Costa, 1983). However, a previous study about the personality questionnaire OPERAS showed that the fact of controlling responses biases in this test, with this procedure, do not affect its convergent validity. Moreover, another study shows that controlling response biases with this procedure do not decrease the convergent and discriminant validity of psychometric measures of aggression (Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco & Lorenzo-Seva, 2012).

Taking all the above into account, the main goal of the current study was to determine the effect of SD and AC on the factor structure of three questionnaires based on the FFM. One of these questionnaires, the Overall Personality Assessment Scale (Vigil-Colet, Morales-Vives, Camps, Tous & Lorenzo-Seva, 2013) was developed using the method proposed by Ferrando et al. (2009). One of the criteria used to choose the items was to minimize their SD and AC levels and remove from the initial pool of items those with high loadings on SD or AC. The other two questionnaires, the FFPI and the BFI were not initially designed with the method proposed by Ferrando et al. (2009), but the effects of SD and AC can also be removed by adding the same four items used in OPERAS as SD markers.

To achieve our goal we assessed the effects of response bias at an EFA level by comparing the simplicity and congruence with the expected FFM of the factor structures of these three questionnaires obtained using the initial and the residual inter-item correlation matrixes (after the SD and AC had been removed). Furthermore, as we have seen above, the effects of acquiescence on the factor structure of questionnaires have usually been found in samples of chil-

dren/adolescents or individuals with a low educational level. In this case, then, we would like to know if response bias also has effects on the factor structure of FFM questionnaires in a sample of adults with a medium-to-high educational level.

Method

Participants

The sample consisted of 392 volunteer undergraduate students on seven different degrees at the Universitat Rovira i Virgili in Spain. Of these, 105 (26.8%) were student teachers (81% women), 50 (12.8%) were social education students (80% women), 38 (9.7%) were pedagogy students (78% women), 94 (24.0%) were engineering students (16% women), 34 (8.7%) were architecture students (50% women), 43 (11.0%) were geography students (22% women) and 28 (7%) were history students (54% women). They were between 18 and 49 years old ($M = 21.5$; $S.D. = 4.2$) and 54% of the participants were females. There are no sex differences in age ($t(336.9) = 1.35$, $p = .18$).

Measures

Big Five Inventory (BFI, Benet-Martínez & John, 1998). It is a 44-item instrument that measures the following factors: Extraversion (EX), Agreeableness (AG), Conscientiousness (CO), Neuroticism (NE) and Openness to Experience (OE). Participants rate each item on a 5-point scale ranging from 1 (disagree strongly) to 5 (agree strongly). The factor structure of the Spanish version is the same as that of the English version, with the following reliabilities: $\alpha = .85$, $\alpha = .66$, $\alpha = .77$, $\alpha = .80$, and $\alpha = .79$ for EX, AG, CO, NE and OE, respectively.

Five-Factor Personality Inventory (FFPI; Hendriks et al., 1999; Hendriks, Kuyper, Offringa & Van der Werf, 2008). We administered the Spanish version adapted by Rodríguez-Fornells, Lorenzo-Seva and Andrés-Pueyo (2001). The inventory consisted of 100 items, twenty for each scale: Extraversion (EX), Agreeableness (AG), Conscientiousness (CO), Emotional Stability (ES), and Autonomy (AU). Participants rate each item on a 5-point scale ranging from 1 (not at all applicable) to 5 (totally applicable). The factor structure of the Spanish version was the same as that reported for the English version, with factor reliabilities of $r_{\theta\theta} = .84$, $r_{\theta\theta} = .84$, $r_{\theta\theta} = .84$, $r_{\theta\theta} = .82$, and $r_{\theta\theta} = .78$ for EX, AG, CO, ES, and AU, respectively.

Overall Personality Assessment Scale (OPERAS; Vigil-Colet, Morales-Vives, Camps, Tous & Lorenzo-Seva, 2013). It is a 40-item instrument (4 markers of SD and 36 content items). Participants must indicate their level of agreement with the item using a five-point scale from "fully disagree" (1) to "fully agree" (5). This test uses the method developed by Ferrando et al. (2009) and gives SD- and AC-free scores for Extraversion (EX), Emotional Stability (ES), Conscientious-

ness (CO), Agreeableness (AG) and Openness to experience (OE) plus SD and AC scores for each individual. It also has good convergent validity with other widely used personality measures. The factor reliabilities for the five scales from which response biases were removed were: $r_{00} = .88$, $r_{00} = .85$, $r_{00} = .77$, $r_{00} = .71$ and $r_{00} = .80$.

Procedure

The tests were administered collectively by a professional psychologist. The participants were volunteers and were asked to answer the three inventories in their classroom. The order in which the questionnaires were administered was randomized for each class. The questionnaires were anonymous, and respondents had to provide only their gender and age.

The data was analysed using MATLAB 2015a (The MathWorks Inc, 2007) and FACTOR 9.02 (Lorenzo-Seva & Ferrando, 2013).

Data analysis

Among the three questionnaires there were a total of 184 missing values, representing the 0.25% of the total data. Taking into account the low number of missing values, we used mode imputation (replacing missing values with the most common values, i.e. the mode value).

We computed the full inter-item correlation matrix and the residual inter-item correlation matrix (removing SD and AC effects) for each test. In all cases we computed polychoric correlations instead of Pearson product moment correlations because they are more suitable for Likert-type items. The four items included in OPERAS as SD markers were also included in the FFPI and BFI.

These matrixes were used to perform EFA and CFA which compared the results obtained using both kinds of data. For the EFA analysis, and for each test, we computed unweighted least squares estimates, and then carried out orthogonal semi-specified Procrustes rotation (Browne, 1972). Therefore, instead of providing a fully specified target, where each value in the target indicates the expected value in the loading matrix after rotation, it was used a partially specified target, where only some values in the target matrix indicate the expected value in the loading matrix after rotation. The specified values are the ones expected to be zero, or as close as possible to zero, in the loading matrix after rotation. It was proposed that the model for each questionnaire should retain five uncorrelated factors, and the ideal pattern matrix was also provided, according to the Five Factor Model. To assess the fit of the rotated loading matrix, the congruence index (Tucker, 1951) was computed between the rotated loading matrix and the ideal loading matrix. Furthermore, to assess the simplicity of the factor structures, we computed Bentler's Simplicity (S) index (Bentler, 1977) and the Loading Simplicity (LS) index (Lorenzo-Seva, 2003). These indexes assess the simplicity of the solution (defined

as the tendency of the items to load on only one dimension): the higher the values the more simple the solution is.

Results

Table 1 shows the congruence coefficients between the factor structures obtained with and without bias and the ideal FFM solution. It should be taken into account that indexes higher than .85 imply a fair congruence between the rotated loading matrix and the ideal loading matrix, while indexes of .95 or higher imply that the rotated loading matrix and the ideal loading matrix are equal (Lorenzo-Seva & Ten Berge, 2006).

Most of the congruence coefficients increased when response biases were removed, and in a few cases they do not increase but remain equal in both factor structures. Furthermore, all the questionnaires showed greater overall congruence when response biases were removed.

Table 1. Tucker's Congruence indexes with and without controlling response bias.

Questionnaire Factor		C.I. with bias	C.I. without bias
BFI	Extraversion	.87	.87
	Agreeableness	.81	.85
	Conscientiousness	.91	.91
	Neuroticism	.92	.93
	Openness	.84	.88
	Overall congruence	.87	.89
FFPI	Extraversion	.80	.83
	Agreeableness	.68	.85
	Conscientiousness	.72	.86
	Emotional stability	.77	.80
	Autonomy	.51	.78
	Overall congruence	.70	.82
OPERAS	Extraversion	.93	.97
	Agreeableness	.88	.89
	Conscientiousness	.90	.90
	Emotional stability	.92	.91
	Openness	.92	.92
	Overall congruence	.91	.91

Note: C.I.: Tucker's Congruence index.

The analyses of the results obtained for each questionnaire show that the congruence values increased slightly for BFI and OPERAS and quite considerably for FFPI. In fact, when response biases were not removed, most of the FFPI congruencies were well below the acceptable values; when they were removed, however, they were fair or close to fair. OPERAS showed the best overall congruence with the theoretical model, and the congruence of BFI was similar.

Table 2 shows Bentler's Simplicity index and the Loading Simplicity index. They indicate the extent to which each item is related to only one dimension. Taking the results for the three questionnaires together, the factor structures obtained when response biases were controlled were simpler. The results show that the simplest solution is the one found in the OPERAS questionnaire, although in this case controlling response biases leads to almost negligible improvement in the simplicity index-

es. BFI showed similar values, although controlling bias slightly improved its factorial simplicity. Again the greatest improvement was found for FFPI.

Table 2. Simplicity indexes with and without controlling SD and AC.

Questionnaire	Without bias		With bias	
	S Index	LS Index	S Index	LS Index
BFI	.93	.42	.89	.40
FFPI	.85	.38	.65	.28
OPERAS	.94	.47	.94	.46

Note: S Index: Bentler's simplicity index; LS Index: Loading simplicity index.

The effects of removing response bias on congruence and simplicity can be illustrated with two examples. Table 3 shows the effect of removing response bias on the factor structure for one of the most affected scales: the Autonomy scale of the

FFPI, the congruence index of which changes from .51 to .78. The congruence index of an almost unaffected scale – the Neuroticism scale of BFI – only increases from .92 to .93.

As can be seen, when response biases were not removed only 11 of the 20 items on the AU scale showed their salient loading on the expected dimension. When response biases were controlled, this number increased to 17. Furthermore, when response biases were removed, secondary loadings were lower. These effects are reflected by the arithmetic mean of the absolute values of the primary and secondary loadings. In this case it can be seen that when response bias were removed the mean loadings on the AU scale increased from $\lambda_m = .31$ to $\lambda_m = .41$ while the mean secondary loadings decreased from $\lambda_m = .18$ to $\lambda_m = .12$: That is, when response biases were controlled the items increased their loadings on the content factor and decreased their cross loadings, which led to a simpler solution.

Table 3. Loadings of the Autonomy scale of FFPI and the Neuroticism scale of BFI on SD, AC, and content dimensions with and without response bias effects (content dimension with the highest loading in bold).

Item	With Bias					Without Bias						
	EX	AG	CO	EE	AU/OP	SD	AC	EX	AG	CO	EE	AU/OP
AUTONOMY (FFPI)												
5	-.19	.42	.04	-.12	.01	-.04	.28	-.13	.12	-.18	.03	.36
10	-.07	.34	.38	-.02	.02	.07	.38	.04	-.03	.08	.17	.47
15	.13	-.10	-.10	.10	-.40	-.11	.43	.02	-.04	-.03	.07	-.48
20	.19	.21	-.25	.01	-.42	-.11	.21	.13	.12	-.27	.04	-.34
25	.00	.28	.28	-.37	.13	.03	.27	.05	.02	.08	-.24	.44
30	-.12	.07	.04	-.67	.19	.06	.30	-.16	.00	.02	-.63	.25
35	.19	.02	.18	.08	-.50	-.10	-.05	.13	-.10	.12	.15	-.37
40	-.21	.38	.21	-.06	.34	-.07	.43	-.14	.18	.06	.03	.58
45	.11	.31	.11	.31	-.37	.12	.14	.04	.23	.08	.36	-.27
50	.23	-.18	-.01	.08	-.66	.05	.20	.16	-.25	-.01	.11	-.61
55	-.16	.44	.18	-.13	.03	.03	.07	-.10	.12	-.04	.03	.39
60	-.33	.21	.13	-.51	.22	.00	.17	-.33	.06	.04	-.43	.38
65	-.31	.34	.19	-.07	.40	.01	.37	-.25	.19	.04	.02	.59
70	-.11	.35	.24	-.08	.17	.14	.22	.03	-.01	-.08	.11	.62
75	.13	.16	.00	.13	-.46	.14	.06	.09	.02	-.08	.20	-.30
80	.18	.11	-.03	.12	-.58	.02	.12	.08	.04	-.01	.16	-.53
85	.25	.14	-.21	.13	-.27	.02	.04	.22	.09	-.22	.15	-.23
90	-.06	.14	.28	-.35	.20	.05	-.06	-.03	-.03	.16	-.26	.39
95	.11	.03	.05	.08	-.46	-.08	.41	.07	-.07	.01	.13	-.35
100	.29	-.04	.10	.12	-.38	-.06	.20	.21	-.02	.14	.12	-.41
NEUROTICISM (BFI)												
4	-.25	-.17	-.12	.52	.04	-.01	-.04	-.25	-.18	-.11	.52	.01
9	-.10	.15	-.05	-.75	.18	.01	.26	-.13	.09	-.03	-.74	.01
15	-.08	-.15	-.01	.67	-.01	-.01	-.05	-.08	-.16	.00	.66	-.03
19	-.01	.15	.00	-.81	.07	.02	.20	-.03	.10	.01	-.81	-.05
26	-.04	.29	.26	.44	.09	-.10	.11	-.06	.27	.25	.45	.04
30	.03	-.19	-.12	.75	.04	.02	-.02	.03	-.21	-.10	.75	.00
35	.04	.26	.09	-.59	.26	-.03	.34	.01	.19	.11	-.58	.06
38	-.04	-.03	-.05	.74	-.07	-.01	.00	-.04	-.07	-.03	.76	-.17
4	-.25	-.17	-.12	.52	.04	-.01	-.04	-.25	-.18	-.11	.52	.01
9	-.10	.15	-.05	-.75	.18	.01	.26	-.13	.09	-.03	-.74	.01
15	-.08	-.15	-.01	.67	-.01	-.01	-.05	-.08	-.16	.00	.66	-.03
19	-.01	.15	.00	-.81	.07	.02	.20	-.03	.10	.01	-.81	-.05
26	-.04	.29	.26	.44	.09	-.10	.11	-.06	.27	.25	.45	.04
30	.03	-.19	-.12	.75	.04	.02	-.02	.03	-.21	-.10	.75	.00
35	.04	.26	.09	-.59	.26	-.03	.34	.01	.19	.11	-.58	.06
38	-.04	-.03	-.05	.74	-.07	-.01	.00	-.04	-.07	-.03	.76	-.17

The same table shows that AU items had low or negligible loadings on SD while the loadings on AC were much greater, and for some items values even reached $\lambda = .40$. In this case, then, it seems that the effects of response bias on the factor structure seem to be mainly due to AC.

In the case of the Neuroticism scale of the BFI, all the items had their salient loadings on the expected dimension, independently of the presence or absence of response biases. The mean loadings on the content dimension were the same for both analyses ($\lambda_m = .66$) while the effect of removing response biases had only a marginal effect, reducing secondary loadings from $\lambda_m = .107$ to $\lambda_m = .091$.

Discussion

The results of the present research showed that response biases have effects on the factor structure of personality questionnaires, even when individuals have a medium-to-high educational level and are not adolescents or children. This effect is probably because response bias introduces sources of shared variance between the items other than the content that they are measuring (i.e. Bäckström, Björklund & Larsson, 2009).

Several authors have reported that response biases, and particularly AC, distort the factor structure of FFM questionnaires, and that the model does not hold when these biases are not controlled. These effects have mainly been found in samples with little education and in children or adolescents (Meisenberg & William, 2008; Rammstedt et al., 2010; Rammstedt & Farmer, 2013; Soto et al., 2008, 2011). In our case, the results reported above seem to show that response biases also affect the factor structure of personality questionnaires based on the FFM when the samples analysed do not consist of individuals with little education or very young individuals, although these effects are smaller than the ones reported for these specific populations.

As we have discussed above, all the questionnaires showed an increase in congruence with the expected factor structure from the FFM perspective, and greater simplicity, but these effects depend on the test. The improvement was greater for FFPI, smaller for BFI and almost unappreciable for OPERAS. It should be taken into account that OPERAS was developed with the loadings on AC and SD of the items: that is, the items chosen were the ones that showed high content loadings and low bias loadings in pilot studies not the items that had high content loadings and high loadings on SD and AC. This process may explain why they changed less after response biases had been removed, which suggests that methods for controlling response bias at the item selection level should be included in the development stage of a personality questionnaire. Therefore, Big Five personality self-reports should include ways to minimize the ef-

fect of these biases in both the questionnaire development stages and in the methods used to analyse the factor structure of data. Moreover, the questionnaire FFPI is considerably longer than OPERAS and BFI: while FFPI has 100 items, OPERAS has 40 and BFI has 44. Therefore, participants take much longer to answer FFPI, so they may get tired or bored, and give more acquiescent and automatic responses. This may be especially true for participants who are not highly motivated to answer the questionnaires, as is the case of the present study done with undergraduate students. However, when participants are highly motivated – for example, in a job selection process – the differences between these questionnaires in terms of acquiescence may be lower.

The method applied in the present research introduces the possibility of analysing not only the effects of AC but also the conjoint effects of AC and SD. Previous research has used a variety of approaches to analyse only the effects of AC but, in our case, the method proposed by Ferrando et al. (2009) allowed us to control both biases, and was used for all the questionnaires. Although it seems that in FFM questionnaires the loadings on AC are much greater than on SD, so AC may be largely responsible for the results discussed above, in other personality questionnaires which assess such undesirable behaviours as aggression and impulsivity, or specific facets of the FFM, SD may play a more important role. Further research, then, should indicate if in these cases SD has a greater effect on the factor structure (Bell & Naugle, 2007; Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco & Lorenzo-Seva, 2012).

This kind of research should be extended not only to other personality measures, but also to other populations, which have higher levels of response biases. We have shown above that very young people or people with little education have high levels of AC, but other populations may also show high levels of response bias. For instance, several studies have found that response biases are highest in the elderly. They show an increase of two standard deviations on SD and one standard deviation on AC in comparison with middle-aged adults (Dijkstra et al., 2001; Stöber, 2001; Soubelet & Salthouse, 2011; Vigil-Colet et al., 2013). As a result it is quite possible that in this population the effects of these biases on the factor structure of questionnaires are greater than the ones reported in the present study. Moreover, further studies should also take into account sex differences, to test the invariance in the factor structures across sex, with and without controlling SD and AC.

Acknowledgments.– This research was supported by a grant from the Spanish Ministry of Education and Science (PSI2014-52884-P), and a grant from the Catalan Ministry of Universities, Research and the Information Society (2014 SGR 73).

References

- Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43, 335-344. doi: 10.1016/j.jrp.2008.12.013
- Bell, K. M., & Naugle, A. E. (2007). Effects of social desirability on students' self-reporting of partner abuse perpetration and victimization. *Violence and Victims*, 22, 243-256. doi: 10.1891/088667007780477348
- 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. doi: 10.1037/0022-3514.75.3.729
- Bentler, P. M. (1977). Factor simplicity index and transformations. *Psychometrika*, 59, 567-579. doi: 10.1007/BF02294054
- Browne, M. (1972). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 115-120. doi: 10.1111/j.2044-8317.1972.tb00482.x
- Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 207-212. doi: 10.1111/j.2044-8317.1972.tb00492.x
- Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, 39(3), 214-227. doi: 10.1037//0003-066X.39.3.214
- Costa, P. T., & McCrae, R. R. (1992). *Revised Neo Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- De Raad, B., & Hendriks, A. A. J. (1997). A psycholexical route to content coverage in personality assessment. *European Journal of Psychological Assessment*, 13, 85-98. doi: 10.1027/1015-5759.13.2.85
- Dijkstra, W., Smit, J. H., & Comijs, H. C. (2001). Using social desirability scales in research among the elderly. *Quality & Quantity* 35, 107-115. doi: 10.1023/A:1004816210439
- Ferrando, P. J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behavioral Research*, 40, 331-349. doi: 10.1207/s15327906mbr4003_3
- 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(2), 364-381. doi: 10.1080/10705510902751374
- 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. 141-165). Beverly Hills: Sage.
- Häggglund, G. (1982). Factor analysis by instrumental variable methods. *Psychometrika*, 47, 209-222.
- Hendriks, A. A. J., Hofstee, W. K. B., & De Raad, B. (1999). The Five-Factor Personality Inventory (FFPI). *Personality and Individual Differences*, 27, 307-325. doi: 10.1016/S0191-8869(98)00245-1
- Hendriks, A. J., Kuyper, H., Offringa, G. J., & Van der Werf, M. P. (2008). Assessing young adolescents' personality with the five-factor personality inventory. *Assessment*, 15(3), 304-316. doi: 10.1177/1073191107313761
- 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. doi: 10.1016/j.paid.2010.04.015
- John, O. P., 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.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: Guilford Press.
- Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, 68(1), 49-60. doi: 10.1007/BF02296652
- Lorenzo-Seva, U., & Ferrando, P. J. (2009). Acquiescent responding in partially balanced multidimensional scales. *British Journal of Mathematical and Statistical Psychology*, 62, 319-326. doi: 10.1348/000711007X265164
- Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2. A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, 37(6), 497-498. doi: 10.1177/0146621613487794
- Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), 57-64. doi: 10.1027/1614-2241.2.2.57
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of consulting and clinical psychology*, 51(6), 882-888. doi: 10.1037/0022-006X.51.6.882
- McCrae, R. R., Herbst, J. H., & Costa, P. T., Jr. (2001). Effects of acquiescence on personality factors structures. In R. Riemann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217-231). Berlin, Germany: Pabst Science Publishers.
- Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44, 1539-1550. doi: 10.1016/j.paid.2008.01.010
- Morales-Vives, F., Camps, E., & Lorenzo-Seva, U. (2013). Development and validation of the Psychological Maturity Assessment Scale (PSYMAS). *European Journal of Psychological Assessment*, 29, 12-18. doi: 10.1027/1015-5759/a000115
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (p.p. 224-239). New York: Guilford 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. doi: 10.1016/j.jrp.2009.10.005
- Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological assessment*, 25, 1137-1145. doi: 10.1037/a0033323
- 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(2), 145-153. doi:10.1027//1015-5759.17.2.145.
- Ruiz, V. M., & Jiménez, J. A. (2004). Estructura de la personalidad: Ortogonalidad versus oblicuidad. *Anales de Psicología*, 20(1), 1-13.
- Ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, 34, 89-102. doi: 10.1207/s15327906mbr3401_4
- 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. doi: 10.1037/0022-3514.94.4.718
- 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. doi: 10.1037/a0021717
- Soubelet, A., & Salthouse, T. A. (2011). Influence of Social Desirability on Age Differences in Self-Reports of Mood and Personality. *Journal of personality*, 79(4), 741-762. doi: 10.1111/j.1467-6494.2011.00700.x
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17) convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222-232. doi: 10.1027//1015-5759.17.3.222
- Ten Berge, J. M. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, 34(1), 89-102. doi: 10.1207/s15327906mbr3401_4
- The MathWorks Inc (2007). *MATLAB - The Language of Technical Computing, Version 7.5*. Natick, Massachusetts: The MathWorks, Inc.
- Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984)*. Washington D.C.: Department of the Army.
- Vigil-Colet, A., Lorenzo-Seva, U., & Morales-Vives, F. (2015). The effects of ageing on self-reported aggression measures are partly explained by response bias. *Psicothema*, 27(3), 209-215. doi: 10.7334/psicothema2015.32

- Vigil-Colet, A., Morales-Vives, F., Camps, E., Tous, J., & Lorenzo-Seva, U. (2013). Development and validation of the overall personality assessment scale (OPERAS). *Psicothema*, 25(1), 100-106. doi: 10.7334/psicothema2011.411
- Vigil-Colet, A., Morales-Vives, F., & Lorenzo-Seva, U. (2013). How social desirability and acquiescence affect the age-personality relationship. *Psicothema*, 25(3), 342-348. doi: 10.7334/psicothema2012.297
- Vigil-Colet, A., Ruiz-Pamies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema*, 24, 310-315.

(Article received: 24-03-2016; revised: 04-11-2016; accepted: 24-11-2016)