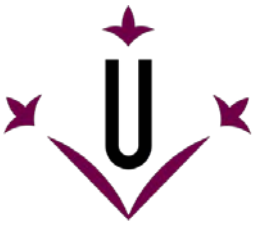

Modeling General, Specific, and Method Variance in Personality Measures: Results for ZKA-PQ and NEO-PI-R

Abstract

Contemporary models of personality assume a hierarchical structure in which broader traits contain narrower traits. Individual differences in response styles also constitute a source of score variance. In this study the bifactor model is applied to separate these sources of variance for personality subscores. The procedure is illustrated using data for two personality inventories - NEO-PI-R and ZKA-PQ. The inclusion of the acquiescence method factor generally improved the fit to acceptable levels for the ZKA-PQ, but not for the NEO-PI-R. This effect was higher in subscales where the number of direct and reverse items is not balanced. Loadings on the specific factors were usually smaller than the loadings on the general factor. In some cases, part of the variance was due to domains different from the main one. This information is of particular interest to researchers as they can identify which subscale scores have more potential to increase predictive validity.

Keywords: Big Five structure, bifactor model, response styles, NEO-PI-R, ZKA-PQ

Modeling General, Specific, and Method Variance in Personality Measures: Results for ZKA-PQ and NEO-PI-R

Most contemporary models of personality assume a hierarchical structure in which broader traits or domains, such as neuroticism, contain narrower traits, such as anxiety or depression (Zuckerman, & Aluja, 2014). These narrower traits are commonly referred to as facets (McCrae & Costa, 2003; Costa & McCrae, 1985), so that facet-analysis refers to the interpretation of a scale in terms of these specific facets. Facet-level analysis is intended to increase predictive validity of personality scores and to provide enriched personality profiles (Paunonen & Asthon, 2001). This is one of the reasons why long, multidimensional questionnaires such as the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), the HEXACO Personality Inventory-Revised (HEXACO-PI-R; Lee & Ashton, 2006), and the Zuckerman–Kuhlman–Aluja Personality Questionnaire (ZKA-PQ; Aluja, Kuhlman, & Zuckerman, 2010), are based on facets. For example, in the case of the NEO-PI-R, the well-known *Big Five* personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) are supposed to account for the common variance of its 240 items, summarized in 30 facets (six facets for each domain).

Although there is some debate about the contribution of narrow measures to predict criteria over and above the broader traits (Salgado, et al., 2014), many studies have shown that narrow measures contribute to the prediction of several outcomes in educational (e.g., McAbee, Oswald, & Connelly, 2014; O'Connor & Paunonen, 2007) and organizational settings (e.g., Barrett, Miguel, Hurd, Lueke, & Tan, 2003; Christiansen & Robie, 2011; Dudley, Orvis, Lebiecki, & Cortina, 2006). In this respect, O'Neill and Paunonen (2013) conclude that the analysis of narrow personality variables provides more information on how

and why personality relates to a wide variety of criteria and enables the researchers to explain the different relationships between certain facets and the criteria.

However, there are also some difficulties in interpreting incremental validity of narrow traits over broader traits. This has to do with how facet scores are computed. In the most common scenario, a raw facet subscale score is obtained as a sum of item responses. When we do this, we cannot differentiate between facets where variance is explained only by the theoretically expected domain (i.e., pure facet scores) and facets where variance is explained by more than one domain (i.e., blended facet scores). This is why several authors (e.g., Anglim & Grant, 2014; Salgado, Moscoso, & Berges, 2013) consider that incremental validity should be shown for the residualized facet scores instead. Likewise, but less acknowledged, the relationship between the residualized facet score and one external criteria might still not be due to the narrow trait but to specific subset of items, or, in the case of self-report external measures, to method artifacts such as the acquiescence bias. Accordingly, Danner, Aichholzer, and Rammstedt (2015) have recently shown that acquiescence bias is moderately stable over time, is consistent across different question types, and can bias the relationship with other variables.

Taking all of the above into account, we propose that the analyses of the internal structure of multidimensional questionnaires through multidimensional bifactor models that also incorporate an acquiescence method factor can be useful to separate several sources of variance (general, specific, and acquiescence bias) for the scores of each residualized facet score. In addition, while the question about the usefulness of narrow traits should be answered empirically for each individual case (Ashton, Paunonen, & Lee, 2014), we suggest here that an upper limit for the magnitude with which each residualized facet score can predict external criteria can be determined by analyzing its specific variance. These proposed analyses are carried out in the present study for two personality inventories: NEO-PI-R and

the ZKA-PQ. In the next two sections we provide details about the bifactor model and the procedures for modeling acquiescence bias.

The Bifactor Model

Application of bifactor models has increased dramatically in the past 10 years. Nowadays, the use of bifactor models is increasingly more habitual in a variety of fields, such as intelligence (Canivez & Watkins, 2010a, 2010b; Gignac & Watkins, 2013), antisocial behavior (Tackett, Daoud, De Bolle, & Burt, 2013), and psychopathy (Patrick, Hicks, Nichol, & Krueger, 2007). In the personality field, Chen, Hayes, Carver, Laurenceau, and Zhang (2012) have illustrated the application of bifactor models to the extraversion domain of the NEO-PI-R in order to test its multifaceted structure.

Bifactor models (see Figure 1C) include a general factor (e.g., broad personality trait) on which all the items load, and several orthogonal specific factors (e.g., narrow personality traits) that represent the common variance that it is not explained by the general factor. Thus, the item common variance is decomposed directly into general and specific common variance. Taking the factor structure of the ZKA-PQ as an example, the variance of the facet scales depends simultaneously on the broader (e.g., neuroticism) and the narrower (e.g., anxiety, depression) latent constructs. We can explore which proportion of the subscale scores is due only to the narrower latent trait using the omega reliability coefficient of the subscale (ω_s; Reise, Bonifay, & Haviland, 2012), which is a reliability estimate for the residualized facet score (i.e., after subtracting the effects of the broader domain factor). Each broad domain scale is assumed to tap a unique broad factor (e.g. neuroticism).

PLEASE INSERT FIGURE 1 ABOUT HERE

Modelling Acquiescence

It is generally agreed that individual differences in the response style are an important nuisance factor that can have systematic effects on the item covariance structure, and can constitute an important source of misfit (e.g., Danner et al., 2015; McCrae, Herbst, & Costa, 2001; Podsakoff, MacKenzie, & Podsakoff, 2012; Rammstedt & Farmer, 2013; Soto, John, Gosling, & Potter, 2008). Thus, it is interesting to find ways of modeling this source of variance so that the model fit can be improved. One of the response styles that can cause distortions in the assessments is acquiescence bias. Acquiescence (disacquiescence) represents the preference for the positive (negative) side of the rating scale (Weijters, Baumgartner, & Schillewaert, 2013). This response style produces inconsistent responses to direct (i.e., positively worded) and reversed (i.e., negatively worded) items. Recently, Savalei and Falk (2014) compared different methods for dealing with acquiescence bias. Savalei and Falk suggest that the best way of modeling acquiescence is by adding a general method factor in the confirmatory factor analysis (CFA) context. The model then incorporates an additional general factor that is orthogonal to all substantive common factors. A person with a high level in this acquiescence method factor is expected to endorse all items no matter what their content. To model this process in a database where items have not been recoded, the factor loadings on this method factor are all fixed to 1. This is equivalent to setting the loadings to 1 for the direct items and to -1 for the reversed items in a recoded database. The parameter of interest to be estimated is the variance of the method factor. This variance is an indicator of the size of the acquiescence bias. It must be noted that in this model all method factor loadings are set to be equal for model identification. In this sense, this model intents to capture an individual's tendency to use the response categories in a consistent manner across items but idiosyncratically among individuals. The parameter recovery of this CFA method has been shown to be robust to the violations of these implicit assumptions (Savalei & Falk, 2014).

In short, the main goal of this study is to decompose personality items variability into its relevant components (i.e., domain, narrow, and acquiescence factors). To do so, a factorial structure must be assumed. This is not an easy task since because, as it occurs with the NEO-PI-R and ZKA-PQ, personality tests are frequently composed of a large number of items and dimensions. For example, the NEO-PI-R has five primary dimensions and thirty narrow factors that are supposed to account for the variance of its 240 items. This large number of items and dimensions preclude modeling all the items in the same analysis. Thus, here we propose two level of analysis. First, each domain will be analyzed separately at the item level. Second, the number of variables will be reduced using item parceling (Marsh, Lüdtke, Nagengast, Moring, & Von Davier, 2013). This allows all domains to be modeled at the same time, which is of particular interest because several facets scores may be interstitial (i.e., influenced by more than one personality domain). If all domains are not modeled within the same analysis, it may happen that the variance attributed to a facet (e.g., impulsiveness) that has a positive secondary loading on a broad domain other than the theoretically expected one (e.g., extraversion) will be overestimated. On the other hand, if two facets within the same broad domain have opposite loadings in another broad domain factor, the correlation between them will be suppressed, and this may also reduce their loading on the intended broad domain factor. Taking these limitations into account, we decided to explore these two levels of analysis. In view of all the above, the specific goals of the present study were: (a) to test whether the inclusion of an acquiescence method factor substantially improves the fit of the model at the item and parcel level, and (b) to assess whether specific latent variance can be reliably measured through facet scores at the item and parcel levels.

Method

Participants and Procedure

García, Escorial, García, Blanch, and Aluja (2012) originally used the data for this study to explore the convergent and discriminant validity of the ZKA-PQ. This database includes the responses of 653 persons (317 men, 336 women) with a mean age of 44.9 years (*SD* = 17.16) to the ZKA-PQ and NEO-PI-R questionnaires. The sample was distributed among the following age ranges: 18–30, 20.7%; 31–40, 19.8%; 41–50, 20.7%; 51–60, 20.2%; and older than 60, 18.7%. Undergraduate students cooperated in the data collection of this study for course credit. The students had instructions to obtain one man and one woman from the following age subgroups: 18–30, 31–40, 41–50, 51–60, and > 60. Subjects were anonymous volunteers recruited among the general population.

Instruments

The ZKA–PQ (Aluja et al., 2010) is composed of 200 items with a 4-point Likert-type response format ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). This instrument measures four facets for each of Zuckerman's alternative five factors of personality: aggressiveness (physical aggression, verbal aggression, anger, and hostility), sensation seeking (thrill and adventure seeking, experience seeking, disinhibition, and boredom susceptibility/impulsivity), activity (work compulsion, general activity, restlessness, and work energy), extraversion (positive emotions, social warmth, exhibitionism, and sociability), and neuroticism (anxiety, depression, dependency, and low self-esteem). In the original study, the authors found a robust 5-factor structure in two Spanish (calibration and validation) and American samples. Factorial congruency coefficients were always higher than .98.

The NEO PI–R (Costa & McCrae, 1992) contains 240 items to which individuals respond on a 5-point Likert-type scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). This instrument was developed to assess the domains of the *Big Five* personality factors: neuroticism (anxiety, angry hostility, depression, self-consciousness, impulsiveness, and vulnerability), extraversion (warmth, gregariousness, assertiveness, activity, excitement,

and positive emotions), openness (fantasy, aesthetics, feelings, actions, ideas, and values), agreeableness (trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness), and conscientiousness (competence, order, dutifulness, achievement, self-discipline, and deliberation).

In a previous study using the same data, Garcia et al. (2012) found that 5 out 20 facet subscales of the ZKA–PQ showed Cronbach's alpha coefficients lower than .70, and the five factor scales had alpha values around .90. In the case of the NEO PI–R, most of the facet subscales (26 out of 30) had reliability coefficients lower than .70, while the five factor scales had values between .83 and .87.

Statistical Analyses

Item-level analysis: modeling each domain separately. Following the theoretical structure of the questionnaires, each broad domain scale is assumed to tap a unique broad factor. A series of seven models were formulated and tested for each personality domain separately: (a) unidimensional model (i.e., all the items load only on a general factor), (b) oblique factor model (i.e., items of each facet load on different correlated specific factors) (c) bifactor model (i.e., each item loads simultaneously on the general factor and its corresponding specific factor). For each of these models, an acquiescence method factor was also tested. For this method factor, loadings of the direct items were fixed to +1, loadings of the reversed items were fixed to -1, and the variance of the factor was estimated. This is the common procedure for testing the effect of acquiescence bias when the reverse items are recoded (Cai, 2010; Maydeu-Olivares & Coffman, 2006; Savalei & Falk, 2014). Finally, a seventh model was tested for each domain where the nonsignificant loadings on the specific factors were set to zero. The ZKA-PQ and NEO-PI-R were analyzed separately. To illustrate these models, the representation for the neuroticism domain of the ZKA-PQ is depicted in Figure 1.

All models were estimated using Mplus 7 (Muthén & Muthén, 1998-2012) and the weighted least squares estimator (WLSMV), which is recommended for categorical data (Muthén, Du Toit, & Spisic, 1997). We used three goodness-of-fit statistics for model evaluation: the chi-square statistic, comparative fit index (CFI), and the root mean square error of approximation (RMSEA). The conventional cutoff values for the CFI are .90 or greater for acceptable fit and .95 or greater for good fit (Hu & Bentler, 1999). RMSEA values between .05 and .08 represent an acceptable fit, whereas values lower than .05 indicate a good fit (McDonald & Ho, 2002). Direct statistical comparisons among nested models[1] were computed. The interpretation of the chi-square difference test was based on the *p*-values of the DIFFTEST option results in Mplus. In addition, nested and nonnested models were compared using the differences in CFI. A difference in CFI of .002 or less is typically adopted as evidence that the imposition of additional constraints does not lead to a significant loss of fit (Meade, Johnson, & Braddy, 2008).

As stated previously, the purpose of the study was to separate the sources of variance for each residualized facet score. That is, to obtain the percentage of variance due to the general factor, the specific factor and the acquiescence method factor. This can easily be done with the application of the bifactor model. The variance of a facet sum score, X_f, can be defined as

$$Var(X_f) = \sum_{m=1}^{M} V_m + V_{AQ} + V_e,$$

where m is a substantive factor (broad or narrow factor), V_m is the variance due to factor m, V_{AQ} is the variance due to the acquiescence (AQ) method factor, and V_e is the unique variance (i.e., uniqueness). Here follows the formulation of each of these elements

[1] The oblique model is not nested in the standard bifactor model where an orthogonally constraint for the specific factors is set (Reise, 2012).

$$V_m = \left(\sum_{\substack{j=1 \\ j \in f}}^{J} \lambda_{jm} \right)^2 Var(F_m)$$

$$V_{AQ} = \left(D - R\right)^2 Var(F_{AQ})$$

$$V_e = \sum_{\substack{j=1 \\ j \in f}}^{J} Var(e_j)$$

In these equations, the sum is made for items included in the facet ($j \in f$), λ_{jm} is the loading of item *j* on the factor *m*, $Var(F_m)$ and $Var(F_{AQ})$ are the variance of *m* and the acquiescence factors, respectively, *D* are *R* the number of direct and reversed items, respectively, and $Var(e_j)$ is the unique variance of item *j*. By dividing each V_m term by $Var(X_f)$ the proportion of variance due to factor *m* can obtained.

Parcel-level analysis: modeling all domains at the same time. Two parcels were constructed for each facet (first and second halves). Then, we applied exploratory structural equation modeling (ESEM; Asparouhov & Muthen, 2009) at the parcel level using the maximum-likelihood estimator of Mplus. Unlike EFA, ESEM models can include both exploratory and confirmatory methods (e.g., equality constraints, fixed parameters, and correlated error terms). In our analyses we defined five correlated ESEM factors corresponding to the five domains (see Figure 2 where the model is depicted). An oblique semi-specified target rotation was defined for these factors (i.e., zero-loadings were defined according with the theoretical model). Additionally, the inclusion of two CFA elements was tested: (a) a specific narrow factor for parcels within the same facet, and (b) an independent acquiescence factor. Regarding the specific narrow factor, loadings of the parcels were constrained to be equal for identification purposes. As in the item-level analysis, loadings on the acquiescence factor were fixed so that they represented the direction of the variables. As far as we know, this is the first time that the acquiescence CFA model is applied at the parcel level. However, its application to item parcels is straightforward. Let's consider a parcel (*P*)

composed of three indicators (two direct items - I_1 and I_2 - and one reversed item - I_3). As in the case of the facet sum score, each item has different contributing sources of variation

$$I_1 = \sum_{m}^{M} \lambda_{1m} F_m + (1) F_{AQ} + e_1$$

$$I_2 = \sum_{m}^{M} \lambda_{2m} F_m + (1) F_{AQ} + e_2$$

$$I_3 = \sum_{m}^{M} \lambda_{3m} F_m + (-1) F_{AQ} + e_3$$

λ_{jm} is the loading on the substantive common factors (F_m), F_{AQ} is the acquiescence method factor, and e are the uniqueness. When these three items are grouped into a parcel, the equation at the parcel (P) level is

$$P = I_1 + I_2 + I_3 = \sum_{m=1}^{M} \left(\sum_{k=1}^{3} \lambda_{km} \right) F_m + (1+1-1) F_{AQ} + \left(\sum_{k=1}^{3} e_k \right),$$

where the loading on the acquiescence method factor is expected to be 2. In this model, variance of facet scores (i.e., $X_f = P_1 + P_2$) can be obtained as

$$Var(X_f) = \sum_{m=1}^{M} V_m + \sum_{m=1}^{M} \sum_{m'=1}^{M} V_{\mathrm{mm'}} + V_{AQ} + V_e$$

As pointed out above, V_m, V_{AQ}, and V_e represent the substantive, acquiescence and unique variances. What is new here is the inclusion of the covariance among the substantive factors (i.e., $V_{mm'}$). All these elements are defined as

$$V_m = \left(\lambda_{1m} + \lambda_{2m} \right)^2 Var(F_m)$$

$$V_{mm'} = \left(\lambda_{1m} + \lambda_{2m} \right) \left(\lambda_{1m'} + \lambda_{2m'} \right) Cov(F_m, F_{m'})$$

$$V_{AQ} = \left(D - R \right)^2 Var(F_{AQ})$$

$$V_e = Var(e_1) + Var(e_2)$$

where λ_{1m} and λ_{2m} are the loadings on the factor m for the two parcels composing each facet,

D is the number of direct items, *R* is the number of reversed items, V_m is the parcel variance due to the factor *m*, $V_{mm'}$ is the parcel variance due to the covariance between factors *m* and *m'*, and V_e is the unique variance. The relative contribution of variance due to these terms can be obtained by dividing each by $Var(X_f)$. There are two points that should be made here. First, perfectly balanced parcels (i.e., the number of positively- and negatively keyed items is the same) have a zero loading on the acquiescence method factor. Secondly, the contribution of the covariance between factors *m* and *m'* can be negative (i.e., if the facet score has loadings of opposite sign on positive correlated factors or if the parcel has loadings of the same sign on negative correlated factors). Finally, an estimation of the facet score reliability can be obtained as (for a general formulation, see Raykov & Marcoulides, 2011, p.180)

$$r_{xx} = \frac{Var(X_f) - V_e}{Var(X_f)}$$

Results

Results will be reported in two sections: one dealing with item-level CFA models for each domain separately and one dealing with parcel-level analyses modeling all domains at the same time. An example of the Mplus syntax for the item-level CFA models as well as full tables for all item loadings for the final item-level CFA models are provided in an online supplement. The rest of the materials can be requested from the corresponding author.

Item-level analysis: modeling each domain separately

Table 1 shows the goodness of fit statistics for all models for the ZKA-PQ. Before including the acquiescence method factor, the RMSEA indexes were close to being acceptable (< .09) for all the models, but the CFI was only acceptable for the bifactor model in the case of the aggressiveness scale (CFI = .92). In almost all the cases, the unidimensional model showed the worst fit and the bifactor model showed the best fit. The only exception was the activity scale, for which the oblique model showed the best fit. The inclusion of the

acquiescence method factor improved the fit of the bifactor models to acceptable levels, except for the activity scale (CFI = .85). Removing the nonsignificant loadings on the specific factors did not worsen the fit values, according to the RMSEA and CFI indexes. Considering the difference in CFI, the bifactor model with the zero-loading constraints produced a better fit than to the rest of the models (ΔCFI < -.002), except for the case of the bifactor model (ΔCFI > -.002). As mentioned above, the only exception was the activity scale. In this sense, the oblique model with and without the acquiescence method factor obtained the best fit (ΔCFI = -.003 and -.027, respectively). Results for the chi-square difference test led to the same conclusions.

PLEASE INSERT TABLE 1 ABOUT HERE

Table 2 shows the average item loadings on the general and specific factors in the final bifactor model. The loadings on the general factor were significant for almost all the items (193 of 200), and the average item loadings varied from .39 (activity) to .47 (aggressiveness). On the contrary, there were 34 out of 200 nonsignificant loadings on the group factors. Average item loadings on the specific factors were usually smaller than the average item loadings on the general factor, except for physical aggression, exhibitionism, restlessness, and work energy. The standardized factor loadings on the acquiescence method factor were significant and varied from .17 (aggressiveness) to .22 (neuroticism).

PLEASE INSERT TABLE 2 ABOUT HERE

For the five domain scales, Table 3 shows that the percentage of variance due to the general factor ranged from 72% (activity) to 85% (sensation seeking). For the five domain

scores, the percentage of common variance explained by the acquiescence method factor was low (smaller than 4%). On the other hand, for the subscales the percentage of variance due to the specific factor was usually lower than what was explained by the domain factor. The highest values were found for exhibitionism (63%), physical aggression (57%), work energy (45%), work compulsion (38%), restlessness (38%), and thrill and adventure seeking (36%). High values mean that these subscales have a high degree of specificity after removing the common variance due to the general factor. The percentage of common variance explained by the acquiescence method factor was usually low (smaller than 3%) except for the boredom susceptibility/impulsivity and the work compulsion subscales (5% and 9%, respectively). Not surprisingly, the number of direct and reversed items is not well balanced in these two subscales.

PLEASE INSERT TABLE 3 ABOUT HERE

Table 4 shows the goodness of fit statistics for all models for the NEO-PI-R. The pattern of results is similar to that obtained for the ZKA-PQ. Before including the acquiescence method factor, the RMSEA indexes were acceptable (< .08) for all the models, but the CFI was unacceptable (CFI < .80). The unidimensional model showed the worst fit and the bifactor model showed the best fit. Again, the inclusion of the acquiescence method factor increased the fit (although not to acceptable levels in this case). CFI values were close to .90 for conscientiousness and openness scales (.88 and 87, respectively), and the worst fit was found for the agreeableness scale (.82). Removing the nonsignificant loadings on the specific factors did not worsen the fit of the bifactor model. When examining the difference in CFI, the pattern of results was similar to the one observed for the ZKA-PQ. In all cases, the bifactor model with the zero-loading constraints obtained a better fit than all the models

(ΔCFI < -.002) with the exception of the bifactor model (ΔCFI > -.002). These results are congruent with the ones obtained for the chi-square difference test. Differences against the final bifactor model with the acquiescence factor were generally non-significant ($p > .05$), with the exception of openness ($\chi^2(12) = 27.2$, $p = .007$) and conscientiousness ($\chi^2(18) = 45.7$, $p = .0003$) scales.

PLEASE INSERT TABLE 4 ABOUT HERE

Table 5 shows the average item loadings on the general and specific factors in the final bifactor model. The loadings on the corresponding broader domain factor were significant for almost all the items (237 of 248), and the average item loadings varied from .31 (extraversion) to .39 (conscientiousness). Regarding the loadings on the specific factors, 62 out of 248 were nonsignificant. Average item loadings on the specific factors were usually smaller than average item loadings on the general factor, except for the extraversion domain (i.e., gregariousness, assertiveness, activity, and excitement seeking subscales) and some subscales in other domains (e.g., impulsiveness, fantasy, values, trust and deliberation). Factor loadings on the acquiescence method factor were significant and varied from .21 (extraversion) to .25 (neuroticism).

PLEASE INSERT TABLE 5 ABOUT HERE

For the five domain scales, Table 6 shows the percentage of variance due to general and narrow factors for scale and subscale scores. The percentage of variance explained by the general factor ranged from 74% (extraversion) to 85% (neuroticism and conscientiousness). The percentage of common variance explained by the acquiescence method factor was low

(smaller than 1%). On the other hand, the percentages of variance due to the specific factors were usually small. The highest values are found for excitement seeking (52%), assertiveness (49%), gregariousness (48%), deliberation (48%), trust (44%), and fantasy (44%) subscales. Despite this, at the same time there were other subscales where the percentage of common variance explained by the specific factor was smaller than 5% (vulnerability, positive emotions, tender-mindedness, and self-discipline). The percentage of common variance explained by the acquiescence method factor was usually low (smaller than 4%). The acquiescence method factor did have some effect in the scales in which only 2 items out of 8 were reversed (depression, warmth, excitement seeking, tender-mindedness, and dutifulness).

PLEASE INSERT TABLE 6 ABOUT HERE

Parcel-level analysis: modeling all domains at the same time

Table 7 shows the goodness of fit for all the tested models for the ZKA-PQ and the NEO-PI-R. In both the cases, the oblique ESEM model does not fit the data satisfactorily but including facet factors or additionally including the acquiescence factor improved the model significantly ($p < .05$ for the chi-square difference test in all cases). The fit indexes for the final models were acceptable for the ZKA-PQ (RMSEA = .05 and CFI = .93) and close to being acceptable for the NEO-PI-R (RMSEA =.04 and CFI =.89).

PLEASE INSERT TABLE 7 ABOUT HERE

For the final models in the ZKA-PQ and NEO-PI-R, the average factor loadings of parcels were higher on their corresponding domain factor than in their specific facet factor (e.g., for the ZKA-PQ, .60 vs. .32; for the NEO-PI-R, were .48 vs. .31). The subscales with

higher factor loadings on the specific factor than in the general factor were almost the same as those found in the item-level analyses (e.g., subscales in the extraversion domain of the ZKA-PQ).

Tables 8 and 9 show the relative contribution of each source of variance to domain and facet scores variance. Regarding the domain scores, the percentage of variance due to the domain latent factor was usually high (e.g., ranging from 71% to 80% for the ZKA-PQ and ranging from 55% to 78% for the NEO-PI-R) and the contribution of facets specificity was low but not negligible (larger than 6% in almost all the scales). Domain scores with a high degree of specificity were activity (in the ZKA-PQ) and extraversion (NEO-PI-R) whereas for neuroticism (for both questionnaires) and conscientiousness (in the NEO-PI-R) the effect of the facets was very small.

PLEASE INSERT TABLE 8 ABOUT HERE

PLEASE INSERT TABLE 9 ABOUT HERE

Regarding the facet scores, the highest specificity values were obtained for psychical aggression (42%), work compulsion (41%), exhibitionism (40%), work energy (28%), thrill and adventure seeking (25%), and experience seeking (17%) of the ZKA-PQ, and for gregariousness (41%), trust (30%), excitement seeking (29%), deliberation (29%), values (28%), and order (26%) of the NEO-PI-R. A low specificity was obtained for depression (0%), hostility (3%), anger (4%), sociability (6%), boredom susceptibility (6%), and disinhibition (7%) of the ZKA-PQ, and for angry/hostility (0%), dutifulness (0%), feelings (1%), self-discipline (1%), vulnerability (3%), and warmth (4%) of the NEO-PI-R. These

results largely replicate those obtained at the item level (the correlation between specificity obtained at the parcel and item levels was .81 and .70, for the ZKA-PQ and the NEO-PI-R, respectively). However, there were also some differences. For some subscales (e.g., angry/hostility, warmth, assertiveness, impulsiveness in the NEO-PI-R) the estimated specificity was smaller when it was computed at the parcel level. This is due to the interstitial nature of these facet scales. An important part of their variance is due to different domains (e.g., angry/hostility, warmth, and assertiveness are also indicators of agreeableness and impulsiveness is also an indicator of extraversion and conscientiousness). When the domains were analyzed separately at the item level the effect of other domains on the facet scores was not tested. For example, at the parcel level, the reliability for angry-hostility (≅ 64) can be decomposed in several sources: (a) related to the broad domain of neuroticism (43), (b) related to broad domains other than neuroticism (17), (c) due to the covariances between neuroticism and the other domains (3), (d) due to the covariances that do not include neuroticsm (-1), (e) related to facet specificity (0), and (f) related to acquiescence factor (1). In this case, subscale loading was positive on neuroticism (.65) and negative on agreeableness (-.41). Given that these loadings are high, the relative contribution of these domains is high. In addition, these two domains are negatively correlated (r = -.09). Thus, there is also an effect due to the covariation between these broad dimensions. This contribution would be small and positive because of this low negative covariation and the fact that the loadings on both broad dimensions are opposite in sign. In comparison to the results for the item-level analysis, we find that the contribution of the specific factor to the variance was non-negligible at the item-level when the neuroticism domain was analyzed separately (28), but this contribution was reduced to 0 at the parcel-level, as this variance was due to other domains. Another divergent result was obtained for some facets (e.g., positive emotions and tender-mindedness in the NEO-PI-R) where specificity was higher according to the new

analyses (4 vs. 19 and 0 vs. 14). Finally, results regarding the percentage of variance due to the acquiescence factor were similar to those obtained at the item level (e.g., work compulsion in the ZKA-PQ and tender-mindedness in the NEO-PI-R showed the largest effect of acquiescence).

Discussion

The current study examined different competing measurement models for personality domains by comparing a bifactor model to unidimensional and oblique factor models. Neither the unidimensional model (assuming only a general factor) nor the oblique factor model (assuming only specific factors) fit the data well. The bifactor model was the best fitting model. However, it should be noted that the fit of the final bifactor model was not good in some cases. This is not uncommon in the confirmatory analysis of personality questionnaires and it is one of the main reasons why it has been argued that the use of CFA to analyze the structure of personality questionnaires at the item level should be limited. Indeed, only short versions such as the NEO-FFI (Costa & McCrae, 1992) have been often analyzed with CFA at the item-level (e.g., Rolland, Parker, and Stumpf, 1998), showing generally poor levels of fit (e.g., Egan, Deary, & Austin, 2000; Gignac, Bates, & Jang, 2007; Panayiotou Kokkinos, & Spanoudis 2004; Schmitz, Hartkamp, Baldini, Rollnik, & Tress, 2001). In fact, McCrae and Costa (2004) revised the item composition of the NEO-FFI in response to this increasing criticism, although this revised version (i.e., NEO-FFI-R) does not clearly improve the psychometric structure of the former one (Aluja, García, Rossier, & García, 2005). However, it should be noted that in these short versions facets are usually ignored, which can partially contribute to the model misfit (Gignac et al., 2007). Thus, some authors (McCrae, Zonderman, Costa, & Bond, 1996) have argued that CFA tests of hypothesized factor structures can be misleading. Hopwood and Donnellan (2010) consider that this bad fit should not be worrying, taking into account the considerable evidence for the criterion-

referenced validity of these questionnaires, nor it should be surprising as it is difficult to write "perfect" items to assess personality. Sources of misfit can be the minor factors (or correlated errors), which easily appear when the content or the phrasing of two items is similar, or might be due to other methodological artifacts (e.g., negatively worded items sharing variance above and beyond the general factor). On the other hand, correcting these sources of misfit can be difficult because additional parameters (e.g., based on modification indexes) might not replicate on cross-validation studies (Hopwood and Donnellan, 2010). Indeed, these authors also propose that broad conventions to provide "thumbs up or down" rules regarding overall model fit should be avoided.

When scales are analyzed with CFA at the facet level, similar difficulties have been found. For example, Hopwood and Donnellan (2010) analyzed seven multiscale instruments, including the NEO-PI-R, with EFA and CFA and found considerable low fit indexes in CFA analysis (e.g., for the NEO-PI-R the CFI index was .61). In this case, two explanations for misfit can be considered: (a) cross-loadings are excluded in CFA whereas some subscales (e.g., impulsiveness) may be interstitial, loading on two or more broad domains, and (b) correlated errors for subscales within a broad domain (e.g., depression and vulnerability) might also share some variance above and beyond the general factor. Accordingly, in this study we found a reasonable fit for the final model, applied at the parcel level, in which cross-loadings were allowed by using ESEM. Fit for the ESEM model might still be improved as facets within a broad domain (e.g., depression and vulnerability) might still share some variance above and beyond the general factor. However, we have avoided these ad-hoc modifications here because they usually do not cross-validate to other samples (e.g., Hopwood and Donnellan, 2010).

Other sources of misfit are individual differences in response style. The two most frequently studied response styles are acquiescence bias and social desirability bias. In the

literature, different strategies have been proposed to deal with these biases and the inclusion of method factors has emerged as an optimal solution. Modeling social desirability is more complex because it requires using a set of social desirability item markers (see, e.g., Ferrando, Lorenzo-Seva, & Chico, 2009). People are more likely to respond in a socially desirable manner in high-stakes assessments. In the present study, the data were collected from the general population in a low-stakes context. Thus, there was not administrated a social desirability questionnaire and we focused on acquiescence bias. We modeled this bias by including an additional general acquiescence method factor in the CFA context following the procedure exposed by Savalei and Falk (2014). In the present work, we have extended this procedure to applications in the ESEM context that are implemented at the parcel level. The inclusion of this factor improved the fit of all the subscales. This result is consistent with other studies that include method factors (e.g., Biderman, Nguyen, Cunningham, & Ghorbani, 2011; Marsh, 1996). Thus, there is increasing evidence that response style produces item covariance that is not explained by the substantive trait factors. This effect is stronger for some scales in which the number of direct and reversed items is not balanced (e.g., tender-mindedness in the NEO-PI-R and work compulsion in the ZKA-PQ). Another important reason for modeling acquiescence is that it can produce bias in the correlation with other self-report criteria (Danner et al., 2015). All things considered, it is crucial to stress again the importance of modeling response styles. Depending on the characteristics of the assessment (e.g., high-stakes vs. low-stakes, motivation), they may have a great impact on the validity of the test scores. Fortunately, new methods and software have been developed for dealing with social desirability and acquiescence style responses (e.g., see Navarro, Vigil Colet, Ferrando, Lorenzo-Seva, 2016).

In this final parcel factor model, we computed the variance due to the narrower traits after the variance in the broad domains was partialized out (i.e., specificities). Results show

that there is a large variability in these specificities, depending on the domain and the questionnaire. These results can be useful for the researchers as it means that the use of some subscale scores (instead of the total score in the domain) has more potential to increase predictive validity. Additionally, for some facets (e.g., angry/hostility, impulsiveness), specificity increases if the variance due to broad domains other than main one is not controlled. Thus, we can hypothesize that incremental validity of these interstitial parcels has room to be larger when other broad domains are not included as predictors. Here follows a summary of the main conclusion for the different domains:

- Regarding the neuroticism domain, there was no facet with high specificity. However, higher potential for incremental validity can be expected for the most interstitial facets, impulsiveness and angry/hostility in the NEO-PI-R, with a large part of variance related to other broad domains (e.g., impulsiveness was positively related to extraversion and negatively with conscientiousness and angry/hostility was negatively related with agreeableness). Interestingly, these facets have previously differentiated from the remaining as representing a subfactor labeled *volatility* externalizing problems of disinhibition, difficulties for controlling impulses and irritability (DeYoung, Quilty & Peterson, 2007).
- For the extraversion domain, many facets were highly specific (e.g., gregariousness), highly interstitial (e.g., warmth, positive emotions), or both (e.g., assertiveness, excitement seeking, exhibitionism). Thus, there is a large potential for incremental validity in this domain, especially when other broad domains are not included as predictors.
- For the conscientiousness domain, large specificities were found only for order and deliberation, the last being highly interstitial. These results suggest that higher incremental validity can be expected from the inclusion of these facets. The

differentiation between these facets and the others is consistent with Roberts, Chernyshenko, Stark, and Goldberg (2005) who, in a factor analysis of a large number of conscientiousness-related subscales, found that order and deliberation loaded both in separated factors (order and self-control), different to the factor in which the remaining NEO measures loaded (industriousness). For the activity domain in the ZKA-PQ, large specificities were found only for work compulsion and work energy, which is highly interstitial, whereas the lowest specificity was found for general activity.

- For the agreeableness domain, several facets were highly specific (e.g., straightforwardness, modesty), highly interstitial (e.g., altruism), or both (e.g., trust). According to DeYoung et al. (2007), altruism and tender-mindedness subscales measure a *compassion* subfactor (i.e., compassionate emotional affiliation with others) whereas straightforwardness, modesty and compliance measure politeness (i.e., consideration of and respect for others' needs and desires). Thus, results regarding the high specificity of straightforwardness and modesty are not expected. DeYoung et al. (2007) and Ashton and Lee (2005) suggest that these facets are markers of honesty/humility dimension in the HEXACO model of personality, which is not well represented in the NEO-PI-R. For the aggressiveness domain in the ZKA-PQ, a large specificity was found for physical aggression and the most interstitial facet was hostility.
- For the openness domain, large specificities were found for fantasy, aesthetics, and values. The most interstitial facet was feelings. Attending to the taxonomy of Woo et al. (2014), fantasy, values and aesthetics would refer to openness to aesthetic, cultural, and self-transformation experiences, whereas ideas is an indicator of intellect, reflecting openness to new intellectual stimulations and should be more connected with plasticity and cognitive behavior. All these facets seem to have some potential, whereas no large

gains of incremental validity are expected with the feelings and actions subscales. In the case sensation seeking domain of the ZKA-PQ, low values of specificities were found for disinhibition and boredom/susceptibility.

The bifactor approach is a useful tool that effectively partials out that common variance (Anglim & Grant, 2014) and provides information about the specific variance of the facet (i.e., the facet score variability that cannot be predicted from the remaining facets in the domain). In short, a low value for specific reliability sets an upper limit for incremental validity. However, it must be noted that the estimation of the variance due to specificity depends on the level of analysis. For example, the percentage of variance due to specificity in the warmth subscale was 4%. This is equivalent to saying that the subscale reliability, that is, the proportion of variance due to the relevant factor, is .04. However, if the analyses were made for each domain separately the percentage of variance due to specificity would be 28% (1% + 23% + 4%), indicating that the subscale reliability is .28. This means that in this case the incremental validity is expected to be larger when the broad domains other than the main domain are not included as predictors.

When presenting these results, it should be noted that the specificity estimates that we obtained are smaller than the indirect estimates obtained by Costa and McCrae (1995). These authors submitted the correlation matrix of the facet scales of the NEO-PI-R to a principal component analysis and then estimated the specificity values as the difference between the alpha coefficient and the communality. They found that 26 out of the 30 scales had specificities greater than .30. In our case, low values of specificity are found for some subscales (e.g., angry/hostility, depression) in comparison to their results. This does not necessarily mean that facets are irrelevant or unreliable. Compared to residualized facet scores, raw facet sum scores are more easily interpretable, reliable, and correlated within the same domain. In this sense, there is some evidence that favors, to some extent, the use of raw

facet sum scores to increase predictive validity. For example, Aluja, Blanch, García, García, and Escorial (2012) found that the average R^2 prediction of personality disorder scale scores increased from .33 to .37 when ZKA-PQ subscales scores were used instead of broad domain scores. In other studies using the NEO-PI-R, Aluja, Cuevas, García, and García (2007) and Dyce and O'Connor (1998) reported the average R^2 in Spanish and American samples, respectively, when predicting personality disorders according to the NEO-PI-R domains and facets. The percentage of variance predicted after including domains was 35%, whereas the prediction after including facets improved that value by 3 and 4% for Spanish and American samples, respectively. Thus, the usefulness of facets is open to debate and, in the fields of personality and work performance there has been much controversy about the incremental validity of facets. Some authors favor the use of facets (e.g., Ashton et al., 2014; Judge et al., 2013; Paunonen, Rothstein, & Jackson, 1999) whereas other authors advise against their use (e.g., Ones & Viswesvaran, 1996; Salgado, et al., 2013). Salgado et al. (2013) showed that a critical point for discussion is to separate common and specific facet variance. In their study, facets of conscientiousness did not predict job performance when the common factor variance (i.e., explained by the general factor of conscientiousness) was excluded. On the other hand, Ashton et al., (2014) found that unique variance of only one theoretically relevant facet (i.e., fairness) did show considerable incremental validity in predicting delinquency beyond and above two broad domains (honesty/humilty and conscientiousness).

Another potential disadvantage of facets is the large number of them that have been defined. As the number of facets included in a criterion validity study increases, the risk of obtaining spurious predictor-criterion relationships also increases (O'Neill et al., 2013). Although the number of facets can be reduced by taking into account the previous research (Dudley et al., 2006; Paunonen & Ashton, 2001), it would raise a range of issues regarding facet selection (Anglim & Grant, 2014). Analyses as those described here can help the

researcher decide which facets can have more potential to increase criterion validity. We have shown that the unreliability of the parcels to measure the specific facets might be a problem for some personality subscales. For example, incremental prediction validity of the facets is typically examined with regression models. Assuming orthogonal predictors, as is the case in the bifactor model approach, the coefficient of determination increase is equal to the squared correlation between the specific factor and the criterion variable. This correlation will be attenuated by the reliability. For example, assuming that the specific factor and the criterion variable correlates .40 and that the reliability for the specific factor scores is .50, this correlation will be reduced to $.40\sqrt{.50} = .28$ and thus the increase would be reduced from $.40^2 = .16$ to $.28^2 = .08$. On the other hand, if one facet has a predictive power beyond and above what is expected by their reliability some concerns can arise regarding the nature of the obtained relation, because it might be attributed to some specific items of the facet and not to the expected construct. In this sense, McCrae (2016) argue that the correlation of outcomes may reflect the effect of the domain (e.g., neuroticism), the narrower trait (e.g., anxiety), or something more specific (e.g., apprehension). In the current work we illustrate how the effect of different sources of variance can be established using factor analysis techniques.

In the current study we showed that the analysis of internal structure provides some insights into the reasons why a subscale might be related to external criteria. In a latent variable approach, reliability, internal structure, and validity are structurally related. In a different view, McCrae, Kurtz, Yamagata, and Terracciano (2011) conclude that internal consistency of scales can be useful as a check on data quality, but appears to be of limited utility for evaluating the potential validity of developed scales (see also McCrae, 2014). More research is needed to test if their conclusions are generalizable to several external criteria and other reliability indexes different from the alpha coefficient, which does not distinguish between different sources of common variance.

Assuming that reliability measures in factor analysis are important for validity, the key question seems to be how to increase the reliability of some subscale scores. One way to do this is to increase the length of the subscales. However, if we do that, the number of facets per domain should be reduced, since otherwise the test would be too long. and it would result in a reduction of the bandwidth of the focus. The bifactor model results can also be useful in guiding the selection of items that increase the reliability of the intended factor (Stucky & Edelen, 2014). For example, items with a higher factor loading on the specific factor relative to the factor loading on the general factor might be better. Another promising approach is to use item response theory to develop a computerized adaptive personality test. In an adaptive test, each person responds to a different set of items (those that are more informative to measure this person's latent trait level). Recently, Makransky, Mortensen, and Glas (2012) have shown that the reliability of the NEO PI-R facets could be substantially improved by applying a multidimensional computerized adaptive testing. For this reason, the use of adaptive tests based on multidimensional models, such as the bifactor model, is a promising area of research (Seo & Weiss, 2015).

References

Aluja, A., Blanch, A., García, L.F., García, O., & Escorial, S. (2012). Zuckerman-Kuhlan-Aluja Personality Questionnaire as a predictor of MCMI-III Personality Disorder Scales: The role of facets. *Personality and Mental Health, 6*(3), 217-227. doi: 10.1002/pmh.1185

Aluja, A., Cuevas, L., García, L.F., & García, O. (2007). Prediction of the MCMI-III personality disorder scores by NEO-PI-R dimensions and facets in Spanish and American samples. *International Journal of Clinical and Health Psychology, 7*(2), 307-321.

Aluja, A., García, O., Rossier, J., & García, L.F. (2005). Comparison of the NEO-FFI, the NEO-FFI-R and an alternative short version of the NEO-PI-R (NEO-60) in Swiss and Spanish samples. *Personality and Individual Differences, 38*, 591-604. doi: 10.1016/j.paid.2004.05.014

Aluja, A., Kuhlman, M., & Zuckerman, M. (2010). Development of the Zuckerman–Kuhlman–Aluja Personality Questionnaire (ZKA–PQ): A factor/facet version of the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ). *Journal of Personality Assessment*, *92*(5), 416-431. doi: 10.1080/00223891.2010.497406

Anglim, J., & Grant, S. L. (2014). Incremental Criterion Prediction of Personality Facets over Factors: Obtaining Unbiased Estimates and Confidence Intervals. *Journal of Research in Personality, 53*, 148-157. doi: 10.1016/j.jrp.2014.10.005

Ashton, M. C. and Lee, K. (2005), Honesty-Humility, the Big Five, and the Five-Factor Model. *Journal of Personality, 73,* 1321–1354. doi: 10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to Salgado, Moscoso and Berges (2013). *Personality and Individual Differences*, *56*, 24-28. doi: 10.1016/j.paid.2013.08.019

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397-438. doi: 10.1080/10705510903008204

Barrett, G. V., Miguel, R. F., Hurd, J. M., Lueke, S. B., & Tan, J. A. (2003). Practical issues in the use of personality tests in police selection. *Public Personnel Management, 32*(4), 497-517. doi: 10.1177/009102600303200403

Biderman, M. D., Nguyen, N. T., Cunningham, C. J. L., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality*, *45*(5), 417–429. doi: 10.1016/j.jrp.2011.05.001

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*(4), 581-612. doi: 10.1007/s11336-010-9178-0

Canivez, G. L., & Watkins, M. W. (2010a). Investigation of the factor structure of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS–IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*(4), 827–836. doi: 10.1037/a0020429

Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher order factor analyses of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, 25, 223-235. doi: 10.1037/a0022046

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219-251. doi: 10.1111/j.1467-6494.2011.00739.x

Christiansen, N. D., & Robie, C. (2011). Further consideration of the use of narrow trait scales. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *43*(3), 183. doi: 10.1037/a0023069

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

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

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21-50. doi: 10.1207/s15327752jpa6401_2

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119-130. doi: 10.1016/j.jrp.2015.05.004

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of personality and social psychology*, *93*(5), 880-896. doi: 10.1037/0022-3514.93.5.880

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, *91*(1), 40. doi: 10.1037/0021-9010.91.1.40

Dyce, J. A., & O'Connor, B. P. (1998). Personality disorders and the five-factor model: A test of facet-level predictions. *Journal of personality disorders*, *12*(1), 31.

694 Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-
695 level analysis suggest N, A and C are more reliable than O and E. *Personality and*
696 *Individual Differences*, *29*, 907-920. doi: 10.1016/S0191-8869(99)00242-1
697 Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure
698 for assessing response bias in questionnaire measures. *Structural Equation*
699 *Modeling*, *16*(2), 364-381.
700 García, L. F., Escorial, S., García, O., Blanch, A., & Aluja, A. (2012). Structural analysis of
701 the facets and domains of the Zuckerman–Kuhlman–Aluja Personality Questionnaire
702 (ZKA–PQ) and the NEO PI–R. *Journal of Personality Assessment*, *94*(2), 156-163.
703 doi: 10.1080/00223891.2011.645935
704 Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-
705 based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*, 639-662.
706 doi: 10.1080/00273171.2013.804398
707 Gignac, G. E., Bates, T. C., & Jang, K. L. (2007). Implications relevant to CFA model misfit,
708 reliability, and the five-factor model as measured by the NEO-FFI. *Personality and*
709 *Individual Differences*, *43*(5), 1051-1062. doi: 10.1016/j.paid.2007.02.024
710 Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality
711 inventories be evaluated?. *Personality and Social Psychology Review*, *14*(3), 332-346.
712 doi: 10.1177/1088868310361240
713 Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure
714 analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*,
715 *6*(1), 1-55. doi: 10.1080/10705519909540118
716 Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013).
717 Hierarchical representations of the five-factor model of personality in predicting job

718 performance: Integrating three organizing frameworks with two theoretical
719 perspectives. *Journal of Applied Psychology, 98,* 875–925. doi: 10.1037/a0033901
720 Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory:
721 two new facet scales and an observer report form. *Psychological assessment, 18*(2),
722 182-191. doi: 10.1037/1040-3590.18.2.182
723 Makransky, G., Mortensen, E. L., & Glas, C. A. (2012). Improving personality facet scores
724 with multidimensional computer adaptive testing: An illustration with the NEO PI-
725 R. *Assessment, 20*(1), 3-13. doi: 10.1177/1073191112437756
726 Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful
727 distinction or artifacts? *Journal of Personality and Social Psychology, 70*(4), 810-819.
728 Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & Von Davier, M. (2013). Why item
729 parcels are (almost) never appropriate: Two wrongs do not make a right—
730 Camouflaging misspecification with item parcels in CFA models. *Psychological*
731 *Methods*, *18*(3), 257-284. doi: 10.1037/a0032773
732 Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis.
733 *Psychological Methods*, *11*, 344-362. doi: 10.1037/1082-989X.11.4.344
734 McAbee, S. T., Oswald, F. L., & Connelly, B. S. (2014). Bifactor models of personality and
735 college student performance: a broad versus narrow view. *European Journal of*
736 *Personality*, *28*(6), 604-619. doi: 10.1002/per.1975
737 McCrae, R. R. (2014). A more nuanced view of reliability specificity in the trait
738 hierarchy. *Personality and Social Psychology Review*, *19*(2), 97-112. doi:
739 10.1177/1088868314541857.
740 McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory*
741 *perspective*. New York: Guilford Press.

McCrae, R. R., Herbst, J. H., & Costa Jr, P. T. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F. Ostendorf, & F. Spinath (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217-231). Berlin: Pabst Science Publishers.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*, 28-50. doi: 10.1177/1088868310366253

McCrae, R. R., Zonderman, A. B., Costa Jr, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552. doi: 10.1037/0022-3514.70.3.552

McCrae, R.R., & Costa, P.T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36,* 587-596. doi: 10.1016/S0191-8869(03)00118-1

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64-82. doi: /10.1037/1082-989X.7.1.64

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568-92. doi: 10.1037/0021-9010.93.3.568

Muthén, B., Du Toit, S. H., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*, *75*, 1-45.

Muthén, L.K., & Muthén, B.O. (1998-2012). *Mplus user's guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Navarro, D. Vigil Colet, A., Ferrando, P.J., Lorenzo-Seva, U. (2016, February). Psychological Test Toolbox. Retrieved from http://psico.fcep.urv.es/utilitats/PsychologicalTestToolbox/index.html

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences, 43*(5), 971-990. doi: 10.1016/j.paid.2007.03.017

O'Neill, T. A., & Paunonen, S. V. (2013). Breadth in personality assessment: Implications for understanding and prediction of work behavior. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 299–332). New York: Routledge.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609–626. doi: 10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K

Panayiotou, G. M., Kokkinos, C., & Spanoudis, G. (2004). Searching for the ''Big Five'' in a Greek context: The NEO-FFI under the microscope. *Personality and Individual Differences, 36,* 1841-1854. doi: 10.1016/j.paid.2003.07.005

Patrick, C. J., Hicks, B. M., Nichol, P. E., & Krueger, R. F. (2007). A bifactor approach to modeling the structure of the Psychopathy Checklist-Revised. *Journal of Personality Disorders, 21*(2), 118–141. doi: 10.1521/pedi.2007.21.2.118

Paunonen, S. V., & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524–539. doi: 10.1037/0022-3514.81.3.524

Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20,* 389-405. doi: 10.1002/(SICI)1099-1379(199905)20:3<389::AID-JOB917>3.0.CO;2-G

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, *63*, 539-569. doi: 10.1146/annurev-psych-120710-100452

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological assessment*, *25*(4), 1137-45. doi: 10.1037/a0033323

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2012). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129-140. doi: 10.1080/00223891.2012.725437

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, *58*, 103–139. doi: 10.1111/j.1744-6570.2005.00301.x

Rolland, J. P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translations of NEO-PI-R and NEO-FFI. *Journal of Personality Assessment*, *71*(2), 269-291. doi: 10.1207/s15327752jpa7102_13

Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, *21*(1), 74-84. doi: 10.1111/ijsa.12018

Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2014). Validity of the five-factor model and their facets: The impact of performance measure

and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology*, 24(3) 325-349. doi: 10.1080/1359432X.2014.903241

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral Research*, *49*(5), 407-424. doi: 10.1080/00273171.2014.931800

Schmitz, N., Hartkamp, N., Baldini, C., Rollnik, J., & Tress, W. (2001). Psychometric properties of the German version of the NEO-FFI in psychosomatic outpatients. *Personality and Individual Differences, 31*, 713-722. doi: 10.1016/S0191-8869(00)00173-2

Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and Psychological Measurement*. doi: 10.1177/0013164415575147

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*(4), 718-737. doi: 10.1037/0022-3514.94.4.718

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S.P. Reise and D.A. Revicki (Eds.), *Handbook of item response theory modeling: applications to typical performance assessment* (pp. 183-206). New York: Routledge/Taylor & Francis Group.

Tackett, J. L., Daoud, S. L. S. B., De Bolle, M., & Burt, S. A. (2013). Is relational aggression part of the externalizing spectrum? A bifactor model of youth antisocial behavior. *Aggressive Behavior*, *39*(2), 149–159. doi: 10.1002/ab.21466

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*, 320-334. doi: 10.1037/a0032121

Woo, S. E., Chernyshenko, O. S., Longley, A., Zhang, Z. X., Chiu, C. Y., & Stark, S. E. (2014). Openness to experience: Its lower level structure, measurement, and cross-cultural equivalence. *Journal of personality assessment*, *96*(1), 29-45. doi: 10.1080/00223891.2013.806328

Zuckerman, M., & Aluja, A. (2014), Measures of Sensation Seeking. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of Personality and Social Psychological Constructs* (pp. 352-380). London: Academic Press.

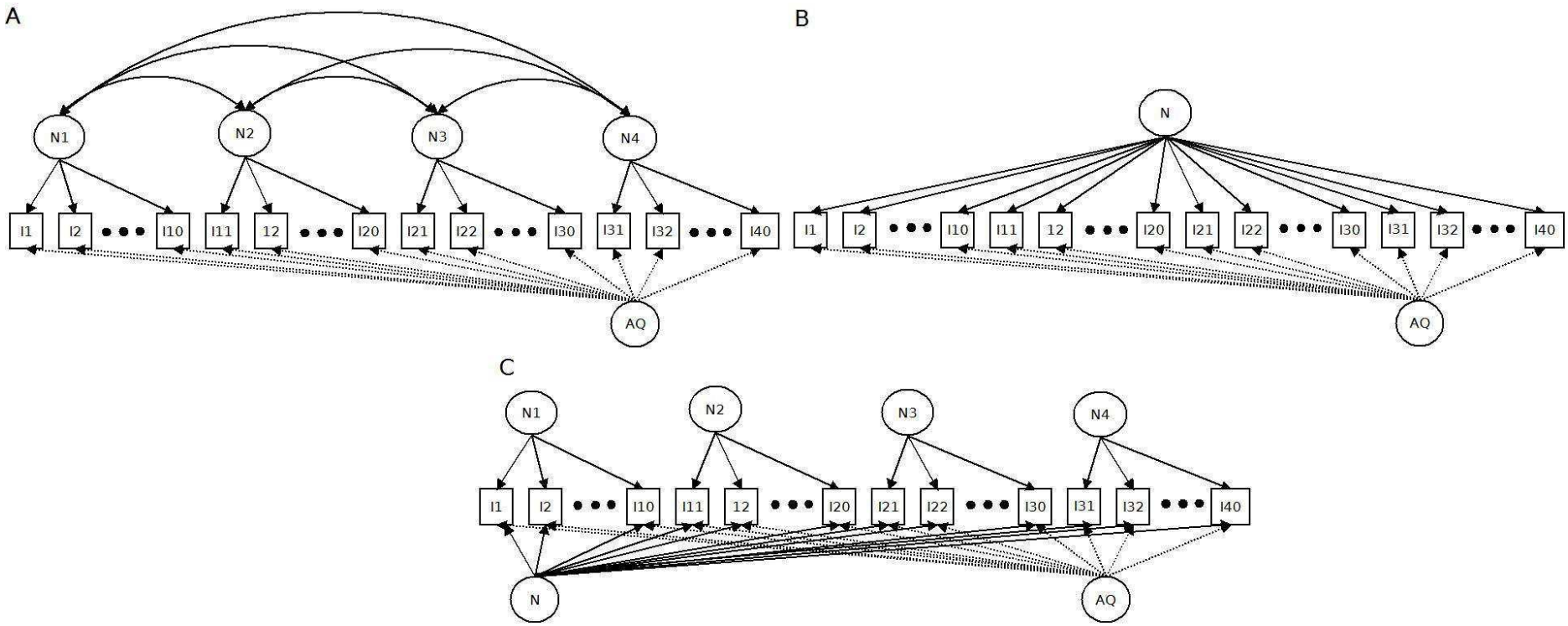

Figure 1. Representation of the different competing models for the neuroticism (N) domain with its four subscales in the ZKA-PQ (anxiety, depression, dependency, and low self-esteem). Model A = Oblique; Model B: Unidimensional; Model C: Bifactor. Each model was tested with and without the acquiescence (AQ) method factor. I1,…, I40 represent the items and N1,…, N4 represent the specific factors,

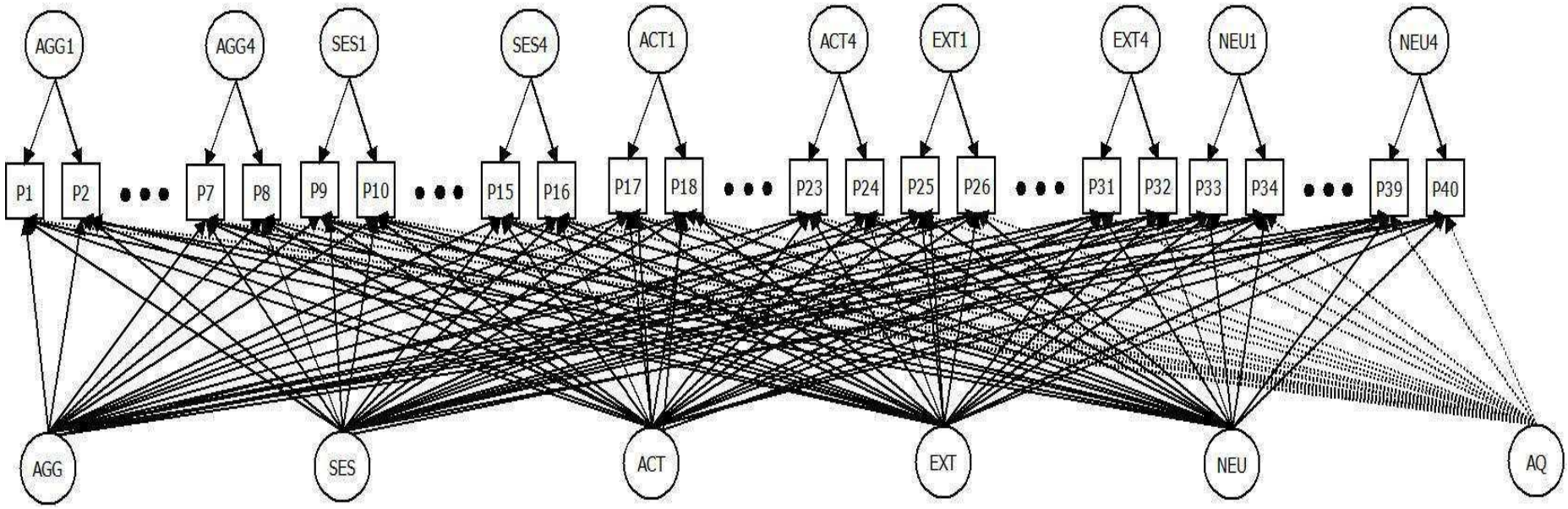

Figure 2. Representation parcel-level factor analysis for ZKA-PQ. Two 5-item parcels are constructed for each independent specific factor (i.e., facets). There are six primary dimensions, namely the correlated big five personality factors and the acquiescence method factor. All these primary dimensions are allowed to load on all items. P1,…, P40 represent the parcels, AGG1,…, NEU4 represent the specific factors, AGG, SES, ACT, EXT, and NEU represent the five domain factors, and AQ represents the acquiescence method factor.

Table 1

Item-level CFA fit statistics for the tested models in the ZKA-PQ

	Without acquiescence factor					With acquiescence factor				
Model	χ^2	*df*	CFI	RMSEA	ΔCFI[b]	χ^2	*df*	CFI	RMSEA	ΔCFI[b]
Aggressiveness										
Unidimensional	4016.1	740	.76	.08	-.174	3752.9	739	.78	.08	-.155
Oblique	2363.6	734	.88	.06	-.055	2091.5	733	**.90**	.05	-.036
Bifactor	1808.4	700	**.92**	.05	-.018	1556.4	699	**.94**	.04	
Final bifactor[a]						1552.5	708	**.94**	.04	**.001**
Sensation Seeking										
Unidimensional	3257.2	740	.77	.07	-.152	2637.7	739	.82	.06	-.095
Oblique	2615.7	734	.83	.06	-.093	2023.8	733	.88	.05	-.038
Bifactor	2028.8	700	.88	.05	-.042	1572.2	699	**.92**	.04	
Final bifactor[a]						1565.4	711	**.92**	.04	**.002**
Activity										
Unidimensional	4414.5	740	.70	.09	-.153	4168.5	739	.72	.08	-.132
Oblique	2522.3	734	.85	.06	.004	2233.3	733	.88	.06	.028
Bifactor	2694.2	700	.84	.07	-.013	2538.5	699	.85	.06	
Final bifactor[a]						2528.8	701	.85	.06	.001
Extraversion										
Unidimensional	4320.7	740	.76	.09	-.164	3788.0	739	.79	.08	-.128
Oblique	3255.3	734	.83	.07	-.092	2829.2	733	.86	.07	-.063
Bifactor	2303.6	700	.89	.06	-.029	1880.2	699	**.92**	.05	
Final bifactor[a]						1880.4	701	**.92**	.05	**.000**
Neuroticism										
Unidimensional	2861.1	740	.81	.07	-.112	2374.8	739	.85	.06	-.068
Oblique	2431.8	734	.85	.06	-.074	1892.6	733	**.90**	.05	-.026
Bifactor	1992.6	700	.88	.05	-.038	1576.3	699	**.92**	.04	
Final bifactor[a]						1567.1	707	**.92**	.04	**.001**

Note. CFI values greater than .90 and ΔCFI values smaller than .002 are shown in bold. [a]: Nonsignificant loadings on the specific factors were fixed to zero. [b]: CFI values are compared to the final bifactor model CFI.

Table 2

ZKA-PQ: number of fixed to zero, nonsignificant or negative item loadings and average item loadings on the general and specific factors in the final item-level bifactor model with acquiescence

Domain / Facets	Number of fixed to zero, nonsignificant or negative item loadings		Average item loadings	
	General factor	Specific factor	General factor	Specific factor
Aggressiveness			.47	
Physical aggression	0	0	.44	.57
Verbal aggression	0	4	.45	.19
Anger	0	2	.60	.21
Hostility	0	3	.40	.20
Sensation Seeking			.44	
Thrill and adventure seeking	0	0	.47	.40
Experience seeking	0	4	.47	.23
Disinhibition	0	5	.50	.12
Boredom susceptibility/impulsivity	0	3	.32	.30
Activity			.39	
Work compulsion	0	1	.38	.38
General activity	0	0	.46	.35
Restlessness	2	1	.27	.30
Work energy	0	0	.44	.45
Extraversion			.45	
Positive emotions	0	0	.48	.35
Social warmth	0	0	.55	.33
Exhibitionism	3	0	.27	.49
Sociability	0	3	.49	.22
Neuroticism			.43	
Anxiety	1	1	.34	.31
Depression	0	4	.46	.18
Dependency	1	2	.38	.24
Low self-esteem	0	1	.53	.30

Note. Each facet contains ten items.

Table 3

ZKA-PQ: percentage of variance due by the general, the specific, and the acquiescence method factors

Domain / Facets	Nº reversed items	% variance due to…		
		general	specific	acquiescence
Aggressiveness	15	83	10	1
Physical aggression	2	33	57	2
Verbal aggression	5	67	12	0
Anger	4	78	10	0
Hostility	4	58	15	0
Sensation Seeking	20	85	9	0
Thrill and adventure seeking	4	51	36	0
Experience seeking	4	66	15	0
Disinhibition	4	75	5	0
Boredom susceptibility/impulsivity	8	38	32	5
Activity	11	72	17	4
Work compulsion	0	37	38	9
General activity	3	53	30	2
Restlessness	4	31	38	1
Work energy	4	42	45	0
Extraversion	19	81	13	0
Positive emotions	5	56	30	0
Social warmth	7	65	23	1
Exhibitionism	3	20	63	2
Sociability	4	69	14	0
Neuroticism	13	82	8	3
Anxiety	4	41	34	1
Depression	3	67	10	2
Dependency	3	51	21	3
Low self-esteem	3	64	21	2

Note. Each facet contains ten items. These values are based on the polychoric correlation matrices.

Table 4

Item-level CFA fit statistics for the tested models in the NEO-PI-R

	Without acquiescence factor					With acquiescence factor				
Model	χ^2	*df*	CFI	RMSEA	ΔCFI[b]	χ^2	*df*	CFI	RMSEA	ΔCFI[b]
Neuroticism										
Unidimensional	4508.1	1080	.64	.07	-.201	3037.3	1079	.80	.05	-.049
Oblique	4250.2	1065	.67	.07	-.176	2732.2	1064	.83	.05	-.018
Bifactor	3941.0	1032	.70	.07	-.147	2526.3	1031	.85	.05	
Final bifactor[a]						2519.6	1039	.85	.05	**.001**
Extraversion										
Unidimensional	4271.1	1080	.63	.07	-.204	3670.6	1079	.70	.06	-.135
Oblique	3270.8	1065	.75	.06	-.090	2630.0	1064	.82	.05	-.017
Bifactor	3083.4	1032	.76	.06	-.073	2449.0	1031	.84	.05	
Final bifactor[a]						2445.5	1033	.84	.05	**.001**
Openness										
Unidimensional	4756.8	1080	.62	.07	-.252	3867.1	1079	.71	.06	-.160
Oblique	4035.7	1065	.69	.07	-.179	3101.3	1064	.79	.05	-.082
Bifactor	3120.2	1032	.78	.06	-.087	2281.8	1031	.87	.04	
Final bifactor[a]						2290.9*	1043	.87	.04	**.000**
Agreeableness										
Unidimensional	4391.0	1080	.60	.07	-.221	3215.6	1079	.74	.06	-.079
Oblique	3884.8	1065	.66	.06	-.161	2721.6	1064	.80	.05	-.021
Bifactor	3621.9	1032	.69	.06	-.133	2516.9	1031	.82	.05	
Final bifactor[a]						2518.4	1037	.82	.05	**.001**
Conscientiousness										
Unidimensional	4652.7	1080	.73	.07	-.154	3606.3	1079	.81	.06	-.075
Oblique	3845.1	1065	.79	.06	-.094	2783.4	1064	.87	.05	-.013
Bifactor	3686.7	1032	.80	.06	-.084	2580.7	1031	.88	.05	
Final bifactor[a]						2597.1	1049	.88	.05	**.000**

Note. CFI values greater than .90 and ΔCFI values smaller than .002 are shown in bold. [a]: Nonsignificant loadings on the specific factors were fixed to zero. [b]: CFI values are compared to the final bifactor model CFI.

Table 5

NEO-PI-R: number of fixed to zero, nonsignificant or negative item loadings and average item loadings on the general and specific factors in the final item-level bifactor model with acquiescence

Domain / Facets	Number of fixed to zero, nonsignificant or negative item loadings		Average item loadings	
	General factor	Specific factor	General factor	Specific factor
Neuroticism			.37	
Anxiety	1	1	.37	.25
Angry/hostility	0	2	.36	.29
Depression	0	1	.46	.22
Self-consciousness	0	3	.30	.18
Impulsiveness	0	1	.26	.33
Vulnerability	0	4	.49	.05
Extraversion			.31	
Warmth	0	0	.38	.32
Gregariousness	1	0	.24	.37
Assertiveness	0	0	.22	.37
Activity	2	1	.27	.30
Excitement seeking	0	1	.22	.40
Positive emotions	0	4	.55	.12
Openness			.35	
Fantasy	0	1	.32	.38
Aesthetics	0	3	.44	.27
Feelings	0	2	.34	.20
Actions	2	4	.27	.13
Ideas	0	4	.49	.25
Values	0	1	.24	.28
Agreeableness			.34	
Trust	0	0	.32	.38
Straightforwardness	2	1	.35	.29
Altruism	0	2	.44	.23
Compliance	0	1	.29	.27
Modesty	0	1	.32	.33
Tender-mindedness	2	5	.30	.02
Conscientiousness			.39	
Competence	0	3	.38	.21
Order	1	2	.27	.30
Dutifulness	0	3	.42	.17
Achievement striving	0	3	.42	.24
Self-discipline	0	8	.53	.00
Deliberation	0	0	.33	.42

Note. Each facet contains eight items. These values are based on the polychoric correlation matrices.

Table 6

NEO-PI-R: percentage of variance due by the general, the specific, and the acquiescence method factors

Domain / Facets	Nº reversed items	% variance due to…		
		general	specific	AQ
Neuroticism	21	85	6	1
Anxiety	4	48	22	0
Angry/hostility	3	43	28	1
Depression	2	59	14	5
Self-Consciousness	3	41	15	2
Impulsiveness	4	25	42	0
Vulnerability	5	73	1	1
Extraversion	19	74	14	1
Warmth	2	42	30	3
Gregariousness	4	21	48	0
Assertiveness	4	17	49	0
Activity	3	28	36	1
Excitement seeking	2	16	52	4
Positive emotions	4	79	4	0
Openness	24	83	8	0
Fantasy	5	31	44	1
Aesthetics	3	56	22	1
Feelings	3	46	16	1
Actions	5	39	09	2
Ideas	3	64	17	1
Values	5	24	32	1
Agreeableness	22	80	9	0
Trust	3	31	44	1
Straightforwardness	5	42	30	1
Altruism	3	57	16	1
Compliance	5	32	29	1
Modesty	4	35	37	0
Tender-Mindedness	2	45	0	7
Conscientiousness	20	85	6	1
Competence	3	52	15	1
Order	5	29	37	1
Dutifulness	2	58	10	5
Achievement striving	3	56	17	1
Self-discipline	4	78	0	0
Deliberation	3	29	48	1

Note. Each facet contains eight items. Percentages based on the polychoric correlation matrices. AQ: Acquiescence

Table 7

Parcel-level CFA fit statistics for the tested models in the ZKA-PQ and NEO-PI-R questionnaires

Questionnaire / Model	χ^2	*df*	CFI	RMSEA	ΔCFI[a]	AIC
ZKA-PQ						
Baseline: correlated domain factors	2687.1	590	.83	.07		120397.8
Baseline + Facets	1806.2	570	**.90**	.06	.068	119556.9
Baseline + Facets + AQ	1437.5	569	**.93**	.05	.029	119190.2
NEO-PI-R						
Baseline: correlated domain factors	4210.6	1480	.81	.05		186593.2
Baseline + Facets	3316.4	1450	.87	.04	.061	185759.1
Baseline + Facets + AQ	2959.5	1449	.89	.04	.025	185404.2

Note. CFI values greater than .90 are shown in bold [a]: Compared with the previous model. AQ: Acquiescence

Table 8

ZKA-PQ: composite reliabilities and variance due to domains, covariances between domains, facets, and acquiescence factors).

		Due to domain…					Due to domains covariances…		Due to facet…				AQ
Domain scores/ Facet scores	r_{xx}	Ag	Ss	Ac	Ex	Ne	main[a]	other[b]	F1	F2	F3	F4	
Aggressiveness	**92**	**79**					4		**6**	1			1
Physical aggression	**85**	**31**	2		4†	3†	3	-2	**42**				2
Verbal aggression	**79**	**68**			5		**-7**	1		**12**			
Anger	**83**	**71**	2†			4	**2**				4		
Hostility	**69**	**40**			**6**†	4	**12**	3				3	
Sensation seeking	**90**		**80**		1		3		4	2	1		
Thrill and adventure seeking	**80**		**52**	2		2†			**25**				
Experience seeking	**77**		**55**		4		-1			**17**			
Disinhibition	**74**		**61**		3		2	-1			**7**		
Boredom susceptibility/ impulsivity	**59**	3	**33**	1†	1†	2	**6**	**3**				**6**	5
Activity	**87**			**71**			1		6	1	1	3	3
Work compulsion	**70**			**24**	1†		-3		**41**				**6**
General activity	**77**	1†	4	**69**	1†		**-6**	-2		**8**			1
Restlessness	**64**	**6**	2	**37**	2	2		2			**13**		
Work energy	**78**		**10**†	**25**	3	3†	**8**	1				**28**	
Extraversion	**89**		1		**73**	1	7		1	2	5	1	
Positive emotions	**78**	1†		4	**39**	**6**†	**17**	3	**8**				
Social warmth	**77**	3†	2†		**56**	1	-1	1		**14**			1
Exhibitionism	**79**	**6**	**7**	1†	**24**		-3	4			**40**		1
Sociability	**67**	1†	4		**53**		5	-1				**6**	
Neuroticism	**89**					**80**	4		1		1	1	2
Anxiety	**69**	**7**		**6**		**37**	4	1	**14**				
Depression	**68**	1			2†	**54**	**9**						1
Dependency	**68**		2†		1	**56**	-4				**11**		2
Low self-esteem	**80**	3†	1	1†	2†	**62**	1	-1				**10**	1

Note. Values higher/lower or equal 6/-6 are shown in bold. [a]: Sum of covariance terms including main domain; [b]: Sum of covariance terms not including main domain, †: Factor subscale loading on the domain was negative. AQ: Acquiescence; Ag: Aggressiveness; Ss: Sensation seeking; Ac: Activity; Ex: Extraversion; Ne: Neuroticism

Table 9

NEO-PI-R: composite reliabilities and proportion of variance due to domains, covariances between domains, facets and acquiescence factors

		Due to domain…					Due to domains covariances…		Due to facet…						AQ
Domain scores/ Facet scores	r_{xx}	Ne	Ex	Op	Ag	Co	main[a]	other[b]	F1	F2	F3	F4	F5	F6	
Neuroticism	**88**	**76**				1	**8**					1	1		1
Anxiety	**63**	**58**			1	4	**-9**		**8**						
Angry/hostility	**64**	**43**			**17**†		3	-1							1
Depression	**70**	**47**	4†		1	1†	**7**				**6**				4
Self-Consciousness	**52**	**25**	4†		1	1†	**6**	1				**13**			1
Impulsiveness	**62**	**20**	**12**		4†	**8**†	4	1					**13**		
Vulnerability	**70**	**45**	2†		2	6†	**10**							3	1
Extraversion	**87**	2	**55**	3			**14**	0		3	2	1	3	2	2
Warmth	**72**	1†	**38**		**23**			0	4						5
Gregariousness	**64**	1†	**18**		4	2†	-2	-1		**41**					
Assertiveness	**68**	4†	**11**	2	**15**†	2	**8**				**25**				
Activity	**60**	2	**29**		1†	**7**	2	-3				**19**			1
Excitement seeking	**71**		**9**	**9**	**6**†	4†	**8**	3					**29**		3
Positive emotions	**74**	4†	**36**	2	1		**11**	1						**19**	
Openness	**88**			**78**			3		2	2		1	2	1	
Fantasy	**68**		3	**23**	1†	**8**†	**10**	1	**22**						1
Aesthetics	**68**			**52**	1		**-6**			**20**					1
Feelings	**56**	**5**	**6**	**35**	3	1	**7**	-4			1				1
Actions	**50**			**29**	1†		2					**16**			2
Ideas	**77**	2†	1†	**59**	1†	1	**-6**						**18**		1
Values	**55**			**23**	3		-2							**28**	1
Agreeableness	**86**		1		**75**	1	1		2	1	1	1	2	1	
Trust	**68**	**6**†	5		**22**			3	**30**						1
Straightforwardness	**62**	1			**33**	3	3			**20**					1
Altruism	**65**		**11**		**37**	3		2			**9**				2
Compliance	**58**	3†	2†		**34**	1†	2	-1				**16**			1
Modesty	**67**	2	2†		**42**		-1						**21**		
Tender-Mindedness	**52**	1	1		**29**	1		-1						**14**	**7**
Conscientiousness	**89**	2			1	**74**	**7**			1				2	1
Competence	**58**	**7**†	2			**28**	**8**	2	**9**						2
Order	**59**	1	1†			**32**	-4	1		**26**					1
Dutifulness	**59**				**9**	**39**	5								**6**
Achievement striving	**69**	1	**7**			**54**	-2					**8**			1
Self-discipline	**71**	1†				**65**	4						1		
Deliberation	**74**	**9**†	**10**†		2	**17**	**6**	-1						**29**	1

Note. Values higher/lower or equal 6/-6 are shown in bold. [a]: Sum of covariance terms including main domain; [b]: Sum of covariance terms not including main domain, †: Factor parcel loadings on the domain were negative. AQ: Acquiescence; Ne: Neuroticism; Ex: Extraversion; Op: Opennes; Ag: Agreeableness; Co: Conscientiousness.